

A New Unified Framework for the Early Detection of the Progression to Diabetic Retinopathy from Fundus Images

Georgios Leontidis*

School of Computer Science, University of Lincoln, Brayford Pool Campus, LN67TS, Lincoln, UK

Abstract

Human retina is a diverse and important tissue, vastly studied for various retinal and other diseases. Diabetic retinopathy (DR), a leading cause of blindness, is one of them. This work proposes a novel and complete framework for the accurate and robust extraction and analysis of a series of retinal vascular geometric features. It focuses on studying the registered bifurcations in successive years of progression from diabetes (no DR) to DR, in order to identify the vascular alterations. Retinal fundus images are utilised, and multiple experimental designs are employed. The framework includes various steps, such as image registration and segmentation, extraction of features, statistical analysis and classification models. Linear mixed models are utilised for making the statistical inferences, alongside the elastic-net logistic regression, boruta algorithm, and regularised random forests for the feature selection and classification phases, in order to evaluate the discriminative potential of the investigated features and also build classification models. A number of geometric features, such as the central retinal artery and vein equivalents, are found to differ significantly across the experiments and also have good discriminative potential. The classification systems yield promising results with the area under the curve values ranging from 0.821 to 0.968, across the four different investigated combinations.

Keywords: Framework; Diabetic Retinopathy; Statistical Analysis; Detection; Classification

1. Introduction

Retinal image analysis remains a challenging and versatile field of study. A number of retinal [1, 2, 3], cardiovascular [4] and systemic diseases [5] can be identified by studying the retinal vasculature and the morphological changes. Although retinal vascular alterations are subtle, early recognition of the signs can be paramount for helping prevent ophthalmologic complications and vision loss. The

*Corresponding author. Tel.: +44(0)7951053731

Email address: gleontidis@lincoln.ac.uk (Georgios Leontidis)

earlier the disease is identified the more efficient the treatment plan can be [6]. In addition to the effort that is placed on providing accurate vessel segmentation and estimating geometric features, which itself entails many challenges, the structure and the dynamic character of the retina is another problem that concerns the present study - especially when it comes to the early detection of the progression from diabetes-NoDR to diabetic retinopathy (DR).

1.1 Background

The human retina is an important, non-invasive window for monitoring the blood vessels [7]. DR is a vision-threatening disease that affects the retinal blood vessels and is primarily a consequence of diabetes [6]. It is characterised by lesions and vascular abnormalities, which include microaneurysms, haemorrhages, cotton wool spots, exudates (bright spots), venous beading, intra-retinal microvascular abnormalities, neovascularisation, loop and fibrous proliferation [8].

To stress how big of a problem diabetes and DR are, from an epidemiological perspective, it is worth quoting the official numbers of the World Health Organization (WHO), which suggests that between 1980 and 2014 the number of people with diabetes has risen from 108 million to a massive 422 million. 2.6 % of global blindness can be attributed to diabetes [6].

The above definition of the problem is enough to appreciate the importance of identifying the signs of the proliferation of diabetes early, at a stage where no lesions exist yet (no DR), but only mild vascular geometric alterations (pre-DR/diabetes-NoDR stage). To do so successfully, three main things need to be achieved. First, the design of comprehensive experiments that focus on the progression of the disease (effects of diabetes to the retina, culminating in first diagnosis of DR), by studying the exact same subjects over a period of time [9]. Second, many different features, and a combination of them, need to be evaluated, not only statistically but also for their discriminative potential. And third, a unified framework is needed that will standardise the process, by combining image processing and machine learning techniques. This will offer a way of studying the progression in successive years within the same cohort, and also accommodate large scale studies in DR and/or other retinal diseases, such as sickle cell retinopathy and retinopathy of prematurity.

1.2 Related Work and Motivation

Numerous studies exist in literature that have presented novelties and advances in retinal image segmentation and registration techniques, as well as in the extraction and estimation of geometric features [10, 11, 12, 13]. All these methods have justified their importance, by stressing their contribution

in studying different retinal diseases, albeit not always happening in practice, as they are not usually made publicly available - or have been developed in a purely research manner.

Regarding the studies about retinal diseases, there are many that have focused on finding associations between vascular changes and retinal diseases, investigating features, such as vessels widths and angles [14, 2], fractal dimension (FD) [15], tortuosity [16], arteriovenous ratio (AVR) [17], haemodynamic features [18, 19], etc. However, no previous studies exist - to the extent the present work does - that investigate at the same time and in multiple experimental designs (for validation and comparison purposes) many candidate features, by not only using proper statistical methods but also building multiple classification models. The former aims to provide information about how each feature changes across the investigated periods of time and the latter about the discriminative potential of them, when utilised for detecting the pre/post- DR stages.

In comparison to other very recent studies that have focused on detecting DR early, the proposed approach is entirely different. For instance, the studies in [20, 21, 22] are utilising lesions, such as microaneurysms and haemorrhages, as features to detect DR. In addition, the aim of the present work supplements the purpose of them, in such a way that we are adding to the existing body of knowledge information about what happens to the retinal vasculature a few years before DR, culminating the diagnosis of DR (onset). As a matter of fact in [21], the No-DR group includes subjects that either have diabetes-NoDR or mild-DR (scale 0 or 1 of the international clinical diabetic retinopathy disease severity scale [23]). In the present study, the mild-DR stage (actually the first year that DR was diagnosed by the DR graders of the hospital) is our terminal group. No other study, to the best of our knowledge, deals with the challenging task of identifying vascular geometric changes well before the onset of DR. In addition, no attempt to automate the DR grading is intended, since this study neither focuses on lesions, nor it targets the same stages of DR, as the studies in [20, 21, 22] do.

Finally, no framework has ever been proposed whatsoever that aims to standardise and facilitate the analyses and classification systems that follow, combining image processing, statistics and machine learning techniques. To accommodate the aims and objectives of this work, the database that is described in the next section and in the supplementary materials was created, by extracting retrospective data.

1.3 Purpose of the Proposed Framework and Studies

The purpose of this work is a) to introduce a novel machine-learning based framework and methodology based on established image processing and machine learning techniques, in order to study the

retinal vascular changes, b) to present statistical analyses tailored to the specific requirements of each experimental design, and c) to demonstrate a comprehensive evaluation of the discriminative potential of the investigated features. The pipeline of the proposed framework is comprised of the following steps, which are going to be described in the next section:

1. **Image registration**
2. **Vessel segmentation and junction measurements**
3. **Automated matching of bifurcations**
4. **Extraction of investigated features**
5. **Statistical analysis and inferences**
6. **Feature selection and classification**

Mathworks Matlab software [24] was utilised for all the algorithms that were used or developed for steps 1-4. Regarding steps 5 and 6, the open source language R and Rstudio [25] were used. The scope of the framework is to accommodate the proposed way of analysis, as introduced in [2, 17], where the same exact segments are studied over a period of time, retrospectively. A total number of forty five geometric features, and derivatives of them, were measured or estimated using the framework, followed by the analysis, feature selection and classification processes.

1.4 Novelty and Contribution

This work presents a number of novelties, including ideas that are introduced for the first time in literature. First, no framework exists that proposes specific steps for the investigation of retinal diseases and the implementation of classification systems. This novel framework, henceforth termed framework for retinal imaging and feature analysis for retinal diseases (FRiAReD), establishes the basis and offers a pipeline that is able not only to automate the process and speed up the analysis of retinal images and the extraction of multiple features, but also ascertain that the bifurcations are correctly registered across the multiple images of the same subject (progression study), and the analysis is valid and therefore meaningful. Second, no other study exists that includes all these features at once and compares many different experimental designs with each other, for exploiting the advantages of matching and studying the same junctions and subjects across a period of time.

The main contributions of this work in the body of literature are summarised in brief as follows:

1. The first major contribution is the proposed framework and experimental design (study 1) that offer an efficient and accurate way to robustly study the alterations in the retinal vasculature,

for identifying and detecting the progression from diabetes-NoDR to DR early. This can have an important clinical implication, because the proliferation of the disease can sometimes be rapid, leading to irreversible damages in the retinal tissue. Coming up with a framework of automated analysis and detection could help clinicians monitor the effects that diabetes has in the retinal vascular geometry more closely and/or adjust the treatment plans of the patients. FRiAReD is a step closer to accomplishing this.

2. The second contribution is that a number of investigated features are evaluated at the same time, and a series of classification models are built that only include geometric features, for testing the potential of deploying into clinical practice such detection systems, which can have important implications. All these solely by including geometric features, without extracting information regarding the lesions.

The rest of the article is organised as follows. Section 2 introduces the methods and studies, followed by section 3 that presents the results and a discussion of the findings. Section 4 is devoted to the conclusion of this article and future work. Section 5 refers to the supplementary materials, which contain important information about the numerical results of the statistical analysis.

2. Methods and Materials

2.1 Data Collection and Studies Overview

All of the studies were performed according to the tenets of the Declaration of Helsinki. In order to investigate and study the progression of the disease over a few year period, and minimise errors coming from factors that cannot be dealt with in advance, such as age, gender, lifestyle and the diversity of different retinas, a retrospective study was planned, as initially described in [2]. To accomplish that, data coming from subjects that progressed from diabetes-NoDR to DR were collected, in order to study the exact same segments over a period of time, and identify vascular changes that one can be confident enough that they can be attributed to the effects that diabetes has to the retinal vasculature. All the images were obtained from the diabetic eye screening service of the Pilgrim hospital in Boston in the UK. The images are graded (R0: Subject with diabetes but no DR, R1: Subject with diabetes with background mild-DR, i.e. only microaneurysms)) by the DR graders of the hospital that the screening takes place, according to the standards and protocols defined by the UK NHS [26]. Images from a total number of 127 subjects were extracted, covering a period of four consecutive years (progressors group - 3 years pre-DR plus the onset of DR). The quality of images is good enough with minimal to

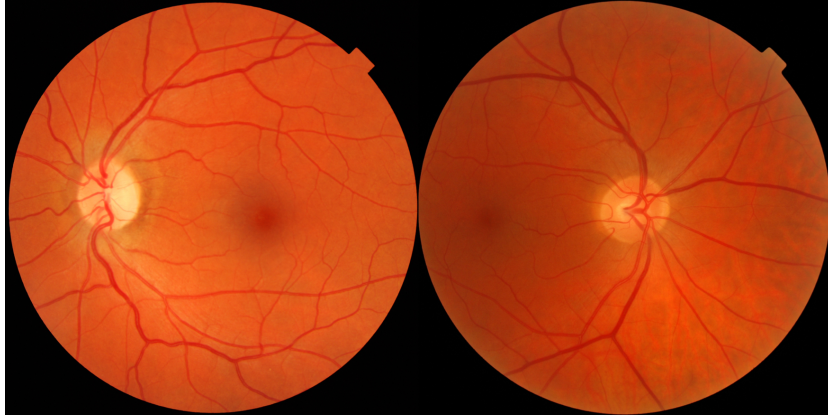


Figure 1: Two good quality and gradeable retinal images can be seen. In both images, all the areas of interest inside the retina can be seen clearly, without artefacts or blurriness.

no artefacts, so that all vessels appear clearly (Figure 1). For all subjects fundus images of both eyes are available. The eye in which DR first appeared was chosen. In the cases where this happened at the same year, one of the two eyes was randomly selected.

In addition, a control group of 27 subjects with diabetes was also extracted that had not progressed to DR at the time of the data collection (non-progressors group). For that group also the last four consecutive years of the disease were selected. This group is needed in order to make some additional analyses, beyond the progression study, for comparison purposes (non-progressors). All subjects in both groups had no other known retinal or cardiovascular disease.

Both the temporal (macula centred) and nasal (optic disc centred) images were extracted, with the latter category to be used for estimating the arteriovenous ratio (AVR) and the central retinal vein and artery equivalents (CRVE and CRAE), and the former for all the other studies. The resolution of these images was 1700X1700 with a field of view 45°. Moreover, the average group age for both the non-progressors and progressors groups were very close (mean age 54.5 (± 7.8) and 53.4 (± 8.4) for progressors and non-progressors, respectively). The extracted data only include images from subjects for whom four consecutive years of progression were available (out of the more than one thousand subjects that were reviewed in the initial database in order to find those with the appropriate history and the availability of all the necessary data).

Table 1 includes information about the planned studies, as will be used for the statistical analysis of all the investigated features across the different stages of both diabetes and DR. The exact data extraction information can be found in the supplementary materials.

Table 1: Summary of the Planned Experiments

Studies	Experimental Design	Description	Particulars
Study 1	Progressors group - four year period of progression with registered segments (four groups - main study)	<ul style="list-style-type: none"> • Four years of progression, starting from the three consecutive years pre-DR until the onset of DR, i.e. three years pre-DR, two years pre-DR, one year pre-DR and onset of DR • Exact same segments and subjects are included in all four years 	Truly identify and associate any vascular changes with the progression of the disease, given that the same exact segments are investigated
Study 2	Progressors group - four year period of progression with non-registered segments (four groups)	<ul style="list-style-type: none"> • Similar to study 1, but the bifurcations are not registered, so no measurements are discarded • Imbalanced groups (non-registered bifurcations; bifurcations might be the same or different across groups) 	Direct comparison with study 1 to investigate whether matching the bifurcations affects the results. This is the standard process of analysis in literature
Study 3	Non-Progressors group - four year period of progression with registered segments without DR (four groups)	<ul style="list-style-type: none"> • Same subjects and segments across a four year period, which does not include DR • Subjects with diabetes who have not progressed to DR 	Evaluate to what extent diabetes affects the retinal vasculature in a four year period, and compare with the progression to DR, as outlined in Study 1
Study 4	Non-Progressors group - four year period of progression with non-registered segments without DR (four groups)	<ul style="list-style-type: none"> • Similar to study 3, but the bifurcations are not registered, so no measurements are discarded • Imbalanced groups (non-registered bifurcations; across groups bifurcations might be the same or different) 	Direct comparison with study 3 to investigate whether matching the bifurcations affects the analysis and the results
Study 5	Averaged non-progressors group (one group) versus averaged progressors group (one group) versus DR-only group (one group)	<ul style="list-style-type: none"> • Does not include the same subjects; the three independent groups of the disease's stage (milestones of progression) are compared 	The purpose is indeed to find out, whether any significant vascular changes occur, during the transition from the non-progressed stage to the progressed, and finally to DR
Study 6	Progressors group - four year period of progression with independent groups (different individuals in each group)	<ul style="list-style-type: none"> • Similar to study 2, this study intends to investigate whether there is any notable reason to prefer the more demanding and sensitive analysis of study 1 • Different subjects and segments across 1'groups 	The stages of the disease are similar to study 1, but the subjects are independent and non-registered across the four year period. Direct comparisons with the results of study 1 will be made, given that the data reflect the same period of progression, but within independent groups

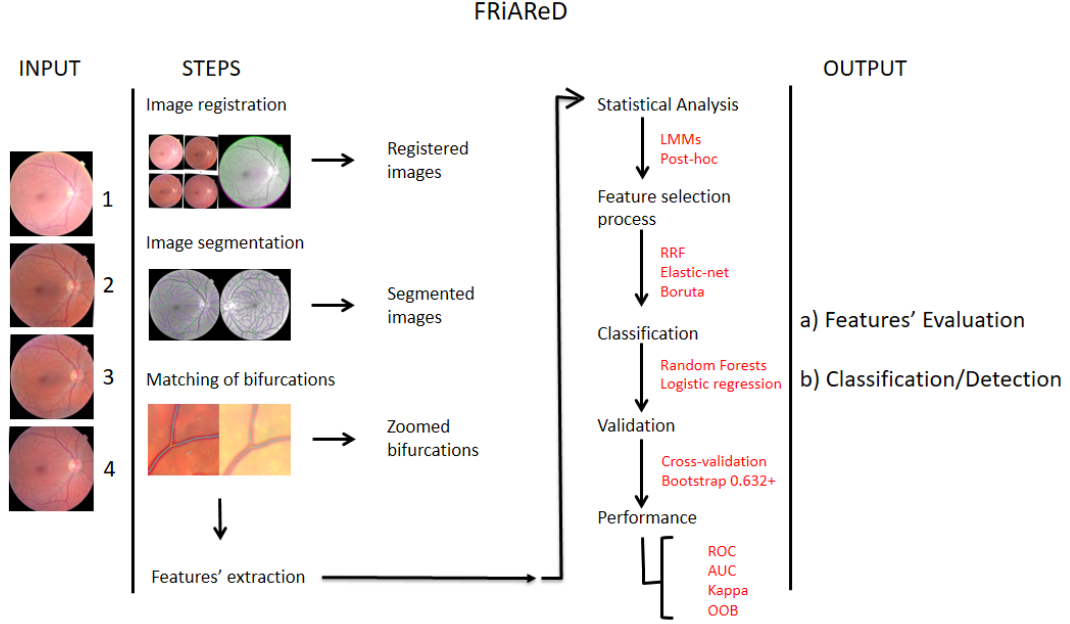


Figure 2: Workflow of the proposed framework

2.2 FRiAReD

An overview of the steps and processes that FRiAReD includes, are illustrated in Figure 2. A detailed description of each step and process is given in this section.

2.2.1 Image Registration

When studying the progression of a disease, it is paramount to be able to compare the same segments over a period of time, mainly for two reasons. First, for eliminating, as far as possible, the errors due to other conditions that affect the measurements, e.g. age, gender, lifestyle, etc., and second for being confident enough that any measured differences can be attributed to the effects of the disease. Doing the opposite, i.e. comparing independent groups and non-registered measurements, requires a much larger sample, because one can inconsistently include and measure large and small vessels across the whole cohort. An efficient and practical way to resolve these issues is to register the bifurcations that come from the same subjects, in order to automatically identify and include only the registered ones, across the images of the same subject.

For the purpose of FRiAReD, the open source library VLFeat [27] was used, implemented in

Matlab, created by Vedaldi and Fulkerson, which offers a complete registration algorithm, based on the scale-invariant features transform (SIFT) and homography estimation with random sample consensus (Ransac). Following the above processes, the images are registered prior to segmenting the vessels, in order the bifurcations to be aligned, and automatically match them, based on their individual coordinates.

In brief, SIFT combines together a feature detector and a feature descriptor [28]. The detector extracts from the image a number of attributed regions in a way that is consistent with some variations of the illumination, the viewpoint and other viewing conditions. The descriptor associates to those regions a signature, which identifies their appearance in a compact and robust way. Then a regular Euclidean distance is used, i.e. the distance between descriptor i from image A and descriptor j from image B [27]. For each descriptor in image A, the algorithm calculates the distance to all the descriptors in image B, matching it with the closest one from B; otherwise the descriptor in A is not matched at all. In order for SIFT to achieve rotation invariance and a high level of efficiency key locations are chosen at maxima and minima of a difference of Gaussian function applied in scale space. A way to efficiently achieve that is to build an image pyramid with resampling between each level [28].

Image gradients and orientations are extracted by processing the smoothed image at each level of the pyramid. At each pixel, $A_{i,j}$, the image gradient magnitude, $M_{i,j}$, and orientation, $O_{i,j}$, are computed using image differences (equations 1 and 2).

$$M_{ij} = \sqrt{(A_{ij} - A_{i+1,j})^2 + (A_{ij} - A_{i,j+1})^2} \quad (1)$$

$$O_{ij} = \text{atan2}(A_{ij} - A_{i+1,j}, A_{i,j+1} - A_{ij}) \quad (2)$$

Following that, a homography is estimated that describes the transformation between the two images (equation 3) [29].

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \Leftrightarrow x_2 = Hx_1 \quad (3)$$

Since the homography is defined up to scale, it has 8 degrees of freedom, hence requiring four point correspondences to estimate [30]. At this point, since not all correspondences are correct, Ransac

is used to find a set of good matches (inliers). This is achieved by randomly choosing 4 matches, estimating the homography by using these matched pairs, and counting how many other matches agree. A match between the point of image A = (X_a, Y_a) and the point of image B = (X_b, Y_b) agrees with the transformation, if the transformed point H(A)=(X'_a, Y'_b) is close to point B [31].

The flowchart of the processes involved can be seen in Figure 3.

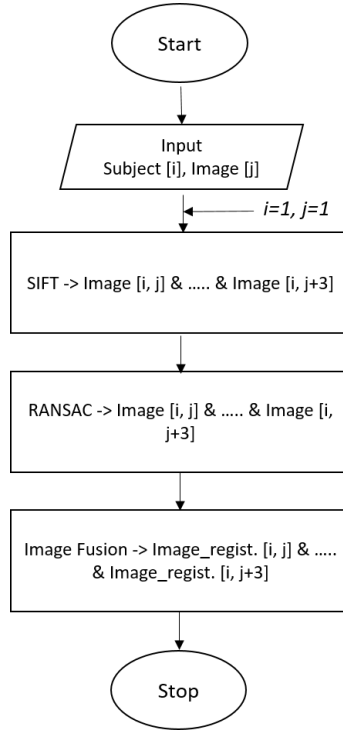


Figure 3: Flowchart of the image registration process.

Figure 4 shows an example of four registered images from our dataset, both separately and fused together.

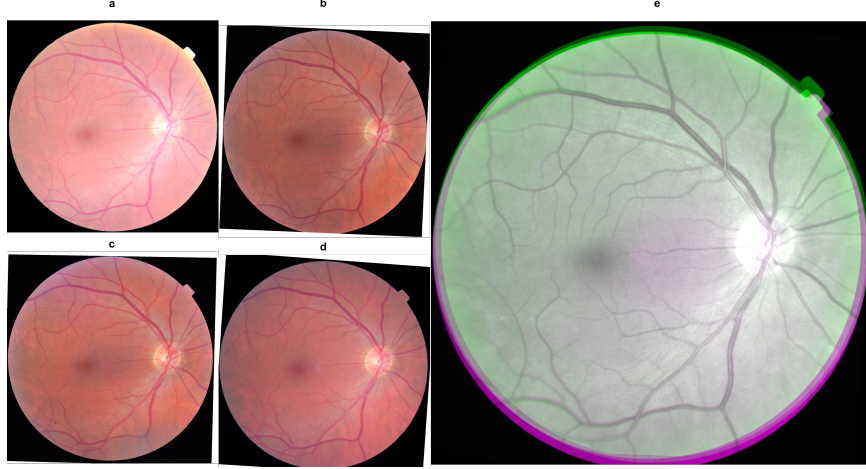


Figure 4: Four registered images and a fused one. a: Reference image. b,c,d: Subsequent registered images. e: Fused image. The bifurcations are adequately aligned, in order to be reliably registered.

2.2.2 Vessel Segmentation and Junction Measurements

Following the image registration, the vessels of the registered images are then segmented, using a segmentation algorithm, whose accuracy, reliability and robustness has been evaluated in literature [32]. This algorithm works by growing a "Ribbon of Twins" active contour model, which uses two pairs of contours to capture each vessel's edge, while maintaining width consistency. At first, the algorithm is initialised using a generalised morphological order filter to identify approximate vessels' centrelines. The output of the filter is defined as follows. Let \mathbf{T} be a binary filter mask of size $M \times M$, with $\#\mathbf{T}$ the number of nonzero elements, and let \mathbf{R} be an $M \times M$ region of the image, centred at a given pixel. Let $S(i, \mathbf{R}, \mathbf{T})$ be the i th largest element of \mathbf{R} masked by \mathbf{T} . Let \mathbf{T}_j^I and \mathbf{T}_j^O be the j th inner and outer tramline filters respectively. Consequently, the filter output is then defined as $\wedge_j(S(\#\mathbf{T}_j^O - 2, \mathbf{R}, \mathbf{T}_j^O) - S(3, \mathbf{R}, \mathbf{T}_j^I) < \theta)$, where $\theta = 0$, is the filtering method. The Ribbon of twins (ROT) is an active contour based model for segmenting retinal vessels. Each ROT has four linked active contours, $\mathbf{v}_c(s, t) = (x_c(s, t), y_c(s, t))$, $c \in \{-2, -1, 1, 2\}$, where x and y are coordinate functions of the parameter $s \in [0, 1]$, and t is time. After identifying the vessel segments, the network topology is determined using an implicit neural function, in order to resolve junction configurations. It also yields precise vessel width measurements, with sub-pixel average width errors. Detailed information and further justification can be found in [32]. Figure 5 shows the flowchart of the steps involved in the segmentation process.

Figure 6 shows an example of the segmentation of two images used in this study. It is worth pointing

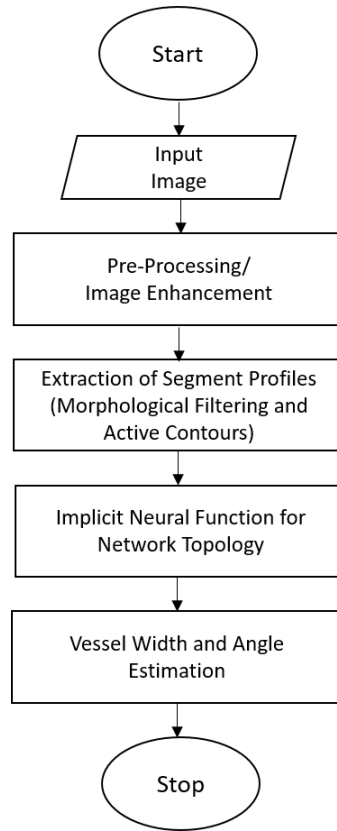


Figure 5: Flowchart of the vessel segmentation process.

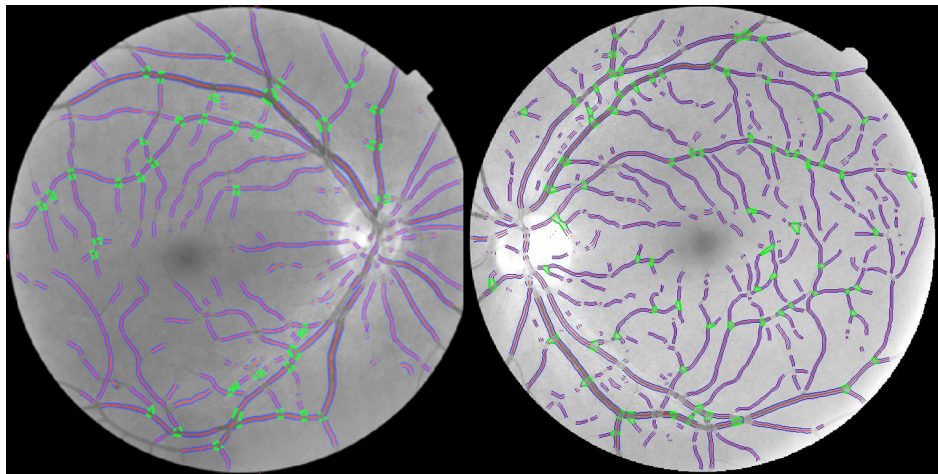


Figure 6: Two images with the vessels segmented can be seen, using the algorithm proposed in [32].

out that all of the specific techniques that are adopted in this study to implement the FRiAReD (SIFT for image registration, active contours for vessel segmentation, etc.) could be replaced with alternative ones, as long as they can accommodate the requirements and the purpose of this framework.

2.2.3 Automated Registering of the Bifurcations

Just like OD and fovea, bifurcations are also anatomical landmarks inside the retina. The position of each bifurcation is more or less fixed, with only negligible changes occurring. Given that the vast majority of the investigated features are extracted in reference to the bifurcations, then a way to make sure that the system will select and compare the same segments during the investigated period of time is by registering the bifurcations. In this way, we make sure that the exact same vessel segments are compared, minimising the chances of making false comparisons. Therefore, although the bifurcations themselves, or more precisely the bifurcation points, do not have any particular importance as features, their almost fixed spatial location, alongside the vessels that form a bifurcation, are highly important and indispensable part of this study. An example of two bifurcations can be seen in Figure 7, which are the same ones, but from two different time periods. Although the registration of the bifurcations is automated, the labelling of arteries and veins is done in semi-automated manner.

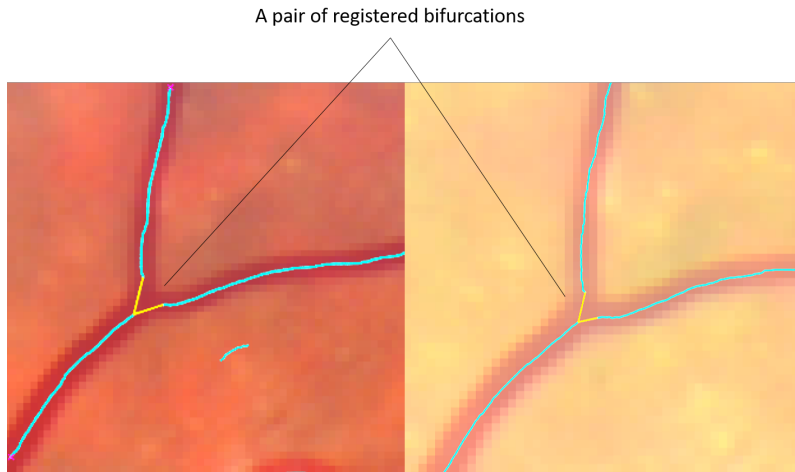


Figure 7: Two bifurcations, belonging to the same subject, but captured in two different time periods of the disease.

Regarding the final iterative process for matching all the segments across the four year period, this is based on multiple conditions, outlined as follows.

1. First, a segment is labelled as artery or vein with a tool presented in [33]. All the labellings were fulfilled by two raters, ending up with a 97,5% kappa inter-rater agreement, thus keeping only

the mutually agreed segments and discarding the rest.

2. Second, the images that come from the same subject are automatically identified, by searching and matching their unique identification numbers, in order to make sure that all four images are available.
3. Following that, only a vein or artery (V or A in our data) is selected at a time; the search is restricted to the respective type of segment, until all of them have been identified.
4. The process starts with the coordinates of a single bifurcation point of an artery and continues with the selection of the first pair of coordinates of the first image of the first subject. The search is only conducted in a 10X10 pixel window about the same coordinates of the other three images of the same subject, which is adequate in order not to miss a not perfectly registered bifurcation - but at the same time not mistakenly include a wrong nearby one. In case more than one bifurcations are found, the algorithm will prompt the user to choose one of them. If both x and y coordinates are within the margins of the defined window in all four images, then the feature name, the period of time, and measurements that correspond to the four registered bifurcations are stored.
5. Finally, the identified matched entries are replaced with zero values and the previous steps are repeated. After the process is finished, the remaining non-zero values are returned, which correspond to the non-registered bifurcations, for possible further investigations. The same process is repeated for the veins.

To make sure that all bifurcations have been registered correctly and the analysis conducted in a proper way, all the registered bifurcations in the corresponding studies were also manually inspected to affirm the suitability of this approach. Given the very good quality of the images and the efficient image registration technique, all the registered bifurcations that were returned after the above process (across the four year period) were identified correctly.

2.2.4 Extraction of the investigated features

As already pointed out, there are numerous studies in literature, having investigated the importance of the retinal geometry in studying different pathologies. However some of the underlying biological processes that might trigger these alterations, are still not well-known [34, 35, 36]. This section provides more information regarding the investigated features and the tools that were used to extract them, prior to describing the experiments and the findings. A cumulative list of all the investigated features,

Table 2: Investigated features as used in the statistical analysis

Features (Short form)	
1. Alpha value of arteries (AlphaA)	24. Venular Angle-to-Branching coefficient ratio(Angle.BC_V)
2. Branching coefficient of arteries (BetaA)	25. Fractal dimension(Fractal)
3. Bifurcation index of arteries (LambdaA)	26. Lacunarity(Lacunarity)
4. Diameter ratio λ_1 of arteries(Lambda1A)	27. Central Retinal Artery Equivalent(CRAE_LEON) [37]
5. Diameter ratio λ_2 of arteries(Lambda2A)	28. Central Retinal Vein Equivalent(CRVE_LEON) [37]
6. Arterial branching angle(ThetaA)	29. Arteriovenous ratio(AVR_LEON) [37]
7. Adjusted Gamma ratio of arteries (artgamma) [37]	30. Central Retinal Artery Equivalent(CRAE_KNUD) [38]
8. Parent's arterial width(Width_parentA)	31. Central Retinal Vein Equivalent(CRVE_KNUD) [38]
9. Large child's arterial width(Width_child1A)	32. Arteriovenous ratio Knudtson(AVR_KNUD) [38]
10. Small child's arterial width(Width_child2A)	33. Average tortuosity [39] (mean_phi)
11. Average junction's arterial widths(Width_allA)	34. Median tortuosity(median_phi)
12. Arterial Angle-to-Branching coefficient ratio(Angle.BC_A) [38]	35. Third quartile tortuosity(75th_phi)
13. Alpha value of veins (AlphaV)	36. Standard deviation of tortuosity(sd_phi)
14. Branching coefficient of veins (BetaV)	37. Average tortuosity [40](mean_tau)
15. Bifurcation index of veins (LambdaV)	38. Median tortuosity(median_tau)
16. Diameter ratio λ_1 of veins(Lambda1V)	39. Third quartile of tortuosity(75th_tau)
17. Diameter ratio λ_2 of veins(Lambda2V)	40. Standard deviation of tortuosity(sd_tau)
18. Branching angle of veins (ThetaV)	41. Average tortuosity[13](mean_psi)
19. Adjusted Gamma ratio of veins (veingamma) [37]	42. Median tortuosity(median_psi)
20. Parent's venular width(Width_parentV)	43. Third quartile tortuosity(75th_psi)
21. Large child's venular width(Width_child1V)	44. Standard deviation tortuosity(sd_psi)
22. Small child's venular width(Width_child2V)	45. Combined tortuosity[41](Tort_blend)
23. Average junction's widths of veins(Width_allV)	

as used in all the experiments of this work, can be seen in Table 2 (as well as the short form of each feature, for visualisation purposes, as they will appear in the models and figures).

2.2.4.1 Geometric Features

Widths & Angles

The vessel widths and branching angles (1-24, Table 2) were estimated using the fully automated active contours algorithm as presented in [32], which is a part of the segmentation algorithm. It provides precise widths and angles measurements with sub-pixel average width errors.

In addition to the above direct measurements, some additional ones have also been proposed in literature and are adopted in this work [42, 43]. These are the bifurcation index λ , the asymmetry ratio α , diameter ratios (λ_1 and λ_2), and the area ratio or branching coefficient β .

Fractal Dimension & Lacunarity

Fractal dimension (FD) and lacunarity (25-26, Table 2) are another two important features that are included in these studies. The former can give us a measure of complexity of a structure, as long as it can be considered a fractal, just like the retinal network [44]. The latter is a measure of heterogeneity of a fractal structure [45].

For the calculation of FD, the well established method of box-counting algorithm (Minkowski - Bouligand dimension) was used [46, 47], whereas the gliding-box algorithm [45] was used for the

estimation of lacunarity.

Arteriovenous ratio

AVR, CRVE and CRAE (27-32, Table 2) were estimated using two different methods. The first one is the widely used method as introduced in [38], and the second one is a more recent and more accurate approach, proposed in [37]. The area of interest, in which all measurements are taken, includes the region where the edges of the vessels course through at 0.5 to 1.0 disc diameters from the OD margin. The vessels are labelled as artery or vein and measured using a semi-automatic tool described in [33].

Tortuosity

Tortuosity (33-45, Table 2) is another very important investigated feature, considered to be among the first indicators of the alterations in the retinal vascular network to appear in retinal pathologies [48]. Tortuosity, in practice, refers to the state or quality of being tortuous, i.e. twisted, having many turns.

In the present studies, three different methods were utilised, described in details in [13, 40, 39]. The image-level tortuosity is expressed in terms of mean, median, standard deviation and third quartile values. In addition, a fourth measure was also used, which is a supervised approach, that provides a tortuosity index capable to reproduce the clinical experts' assessments [41]. This index is a linear, weighted combination of the other three.

2.2.5 Statistical Analysis and Inferences

Each feature under investigation was evaluated using Linear Mixed Models (LMMs). Mixed models refer to a variety of models that have as main variables both fixed and random effects, which are combined linearly with an error term to determine a response variable. LMMs capture individual differences by estimating a different random intercept for each subject [49, 50].

For each feature an LMM is constructed, with the feature as the response variable, the year of measurement (disease stage) as the fixed effect and the subject identity as the random effect. A second, so called "restricted or null model" is also constructed for each feature, with the subject identity as the sole independent variable. The two models are fitted using maximum likelihood, p-values are calculated from the likelihood ratio X^2 (chi-squared) test of the log-likelihoods of the two models, with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) estimated as well. The inferences are made by comparing the two models with each other in order to investigate whether or not the progression of the disease significantly affects the features. The restricted model was also used to estimate the intra-class correlation (ICC), which describes the proportion of the

total variance that can be explained by the individual subject identity. The Tukey-test was used to make post-hoc comparisons and investigate the relationship between the year of measurement and the statistically significant features. All the relevant assumptions that need to be met, e.g. normality of the residuals, were tested either by visually inspecting the data (Q-Q plots) or by corresponding tests (Shapiro-Wilk), prior to running any analysis.

As mentioned, AIC, BIC and p-values are reported. P-values have undergone a great deal of controversy in whether or not they are a suitable approach of evaluating an effect and communicating results [51]. In literature, particularly for LMMs, it is suggested to be avoided and instead use other metrics, such as AIC and BIC. In addition, in LMMs the p-values cannot be directly extracted, because the estimation of the denominator’s degrees of freedom is not straightforward [52].

In the present study all these metrics are going to be communicated in order to give an overall performance indication and avoid only to follow a specific evaluation metric. Moreover, for the reasons elaborated and explained in details in [53], regarding whether the p-values need to be adjusted in multiple comparisons or not (classicists vs rationalists), and also the fact that hierarchical models in general might not need such approaches [54], the reported p-values are not adjusted. Besides, given that all the information needed to adjust the p-values is reported in this work (appendix A in supplementary materials), one is able to do so, if needed. The significant features will be highlighted, with the threshold being p-value less than or equal to 0.05.

A metric similar to the R^2 of the ordinary linear regression, the Ω_0^2 , is also reported. It was proposed for the LMMs in [55], as a generalised metric for estimating the explained variance or the effect size. It can be estimated by

$$\Omega_0^2 = 1 - \frac{\sigma^2}{\sigma_0^2} \quad (4)$$

where σ^2 is the residual variance of the full model, i.e. the amount of variation in the response Y that is not explained by the covariates, and σ_0^2 is the residual variance of an intercept-only null model.

2.2.6 Feature Selection and Classification

In order to test the discriminative power of the proposed features, two different approaches were followed, given their popularity and established performance in a wide range of applications. Both of these algorithms are selecting a subset of features, based on a fine tuning of the penalty parameters, prior to the classification process.

First, a regularised random forests classifier (RRF) was used, which is an adjusted for the feature

selection process random forests classifier, as proposed in [56]. In short, the RRF can select a compact feature subset, by including an additional penalty coefficient, creating a regularised information gain,

$$Gain_R(X_i, u) = \begin{cases} \lambda \cdot Gain(X_i, u) & i \notin F \\ Gain(X_i, u) & i \in F \end{cases} \quad (5)$$

in which u refers to the node and F is the set of indices of features used for splitting in the previous nodes. The parameter $\lambda \in (0,1]$ is the penalty coefficient.

Second, an elastic-net logistic regression was also used, which includes a feature selection step, based on the least absolute shrinkage and selection operator (Lasso) and ridge regression, creating a hybrid penalty (trade-off) for the coefficients of the features (L1- and L2- norms) [57].

In addition to the above two methods that will be used both as classifiers and feature selection processes, a third feature selection process was also utilised. Because the most challenging and important input for a classification system are the features with the highest discriminative potential, a third feature selection-only method is used, the all-relevant feature selection algorithm (boruta) [58]. In brief, it works by creating random features (shadowed attributes), which are then compared with the real features, based on a Z-score. Given that it is a wrapper around random forests, it can provide a numerical ranking of the features as an importance measure. Then the maximum score among the shadowed attributed (MSA) is found (in the corresponding figures 7 and 9, this is referred as "shadowmax"), which is compared with the corresponding scores of the other attributes. The attributes that are found to score statistically significantly higher than the MSA ones are marked as important (selected features).

2.2.6.1 Validation and Metrics

In order to evaluate the performance and validate the classification models, a number of approaches were adopted. The evaluation process was conducted by using a 1000 repeated 10-fold cross validation [59] and a 0.632+ bootstrap [60]. The different models were compared according to their performance in terms of the average area under the receiver operating characteristic curve (AUROC or AUC), but additionally, other metrics are also utilised and/or reported, such as kappa and out of bag error (not for the model comparison, but as a complementary to AUC metric; only for the random forests). Kappa is preferred to the conventional accuracy, because the latter can be a misleading metric of performance if used alone, especially when comparing the performance of different classifiers [61]. That is because

Table 3: Summary of the classifiers, feature selection process and main evaluation metrics

Techniques		Overview	Process
Logistic regression		Linear classifier	Classification
Random forests		Non-linear classifier	Classification
Elastic-net		Regularisation & mixing parameters for LASSO and Ridge regression	Feature Selection
Regularised forests	random	Regularised information gain using penalty parameter	Feature selection
AUC		Represents a probability, the closer to 1 the better	Evaluation metric
ROC		Graphical plot for the performance of the classifiers	Evaluation metric
kappa		Adjusted accuracy based on the balance of the classes	Evaluation metric
OOB		Prediction error using bootstrap aggregating	Evaluation metric

a 75% of accuracy in a balanced design (50-50) is not the same as a 75% of accuracy in an imbalanced design, where the largest class represents the 70% of the whole sample, for instance. Therefore, there is another metric, known as kappa, which compares the observed accuracy with the expected accuracy, yielding a metric that can be used for evaluating both the performance of a single classifier and also across classifiers [62]. Table 3 recaps the approaches and the evaluation metrics that are going to be reported.

2.2.6.2 Classification and Training Process

The classification process was conducted for the following four combinations:

1. Comparison of images (averaged) obtained during three years before DR onset vs. an image at DR onset in the same subjects (DR progressors)
2. Comparison of images (averaged) obtained during four years in subjects who did not developed DR (non-progressors) vs. image at DR onset in DR progressors
3. Comparison of images (averaged) obtained during four years in subjects who did not developed DR (non-progressors) vs. images (averaged) obtained during three years before DR onset
4. Comparison of images (averaged) obtained during four years in subjects who did not developed DR (non-progressors) vs. images (averaged) obtained during three years before DR onset vs. an image at DR onset (DR progressors) - One versus All method

Combination 1 includes the average of the three year period before the onset of DR, which offers a more realistic and representative sample of the diabetes-NoDR condition versus DR. **Combination**

2 is similar to **combination 1**, but during this time, the group includes a non-progressors cohort of subjects with diabetes, for which the last four available consecutive years of diabetes-NoDR are combined and averaged, in order to form this group that will be tested against the DR one. In practice, combination 1 includes data from subjects that is known that they are progressors (features are averaged across the three year period, namely three, two and one year pre-DR), whereas the non-progressors group includes subjects with diabetes that have not progressed to DR (i.e. they are at least in earlier stage than the progressors group). **Combination 3** investigates whether two groups of subjects with diabetes that at first sight belong to the same group but in reality they come from predefined periods, can be correctly identified. In a prospective study this information is not available and therefore subjects from both categories are mixed up together in the sample. In our case, the images have been collected in a retrospective way, so it is known a priori that the subjects in the first group have not progressed to DR yet. Therefore, the purpose here, is to investigate whether two such different cohorts are being (have been) affected by diabetes the same or to a different level that actually defines the progression to DR, and can be considered as such. Finally, **combination 4**, in contrast to combinations 1 and 2, provides an overview of whether the three main groups, i.e. subjects with diabetes but No DR (non-progressors - control group), subjects with diabetes but No DR (progressors group), and DR subjects, can be efficiently differentiated among each other in a joint system. As it will be stressed further later, if this combination yields a good performance, it can possibly introduce an alternative way of studying the progression in a clinical environment, since the first two categories, i.e. subjects with diabetes but No DR (non-progressors - control group) and subjects with diabetes but No DR (progressors group), denote the progression of the disease.

As aforementioned, the primary objective of this work is to evaluate the discriminative power of the entire feature set, as an additional, complementary to the statistical analysis step. The secondary objective is to find the optimum model for each of the above combinations, in order to assess the potential of these classifiers to having a possible clinical interest. All the features were normalised before any of the processes is executed. All of the combinations contain balanced classes. An exemption is the one vs all process, where although the three classes are still balanced, two classes are merged each time, creating a new one, and therefore performing the classification versus the third class. The relevant lists of features are provided in Tables 2 and 4. In comparison to the features in Table 2, which were used for the statistical analysis, for the classification the majority of them is the same, except for some differences that exist, hence Table 4. The differences refer to the widths and angles of the

vessels, which are summarised based on their descriptive statistics, in order to have equal observations for each feature in the classification, and also one observation per image.

Table 4: Additional features for the classification

Features (Short form)	
1.Arterial average branching coefficient(BC_Mean_Art)	2.Standard deviation of arterial widths (arterySD)
3.Average Branching coefficient of veins (VeinBc)	4.Standard deviation of venular widths (Vein_SD)
5.Average arterial widths (MeanArt)	6.Average venular widths (MeanVein)
7.Median arterial widths (MedArt)	8.Median venular widths (MedVein)
9.Standard deviation of arterial angles (ArtAngleSD)	10.Standard deviation of venular angles (VeinAnglSD)
11.Average arterial angles (MeanAngArt)	12.Average venular angles (MeanAngVein)
13.Median arterial angles (MedAngArt)	14.Median venular angles (MedAngVein)

The **feature selection** process includes various techniques (Section 2.2.6) and steps that are described below. For the combinations 1-3 the steps are exactly the same, but for combination 4 the process is slightly different. Everything was fulfilled using Rstudio and a number of packages [25, 63, 64, 57, 56, 65].

As was previously described, each of the feature selection algorithms follows a specific process for producing a feature subset. Given the randomness behind all the processes involved, the feature subsets slightly change every time the algorithms are executed. Initially, the data are randomly partitioned into 10 balanced folds. Leaving one fold aside, the feature selection process is executed 10 times in total, ending up with 30 different models (10 from each selection process). Each of these models inevitably includes a slightly different feature subset. In order to arrive at this point and find the optimum subset, the whole feature set is penalised by the algorithms. Given the complexity of the feature selection processes, in order to find the optimum regularisation parameters that give the final subset for each fold, the 0.632+ bootstrap process is executed with 10000 iterations.

Regarding the elastic-net, the parameters of α and λ were bootstrapped within the range of 0 and 1, with a step of 0.01. Similarly, the regularised random forests process was bootstrapped for the parameters mtry, i.e. the square root of the number of predictors sampled for splitting at each node rounded down (4 and 5; 6 and onwards gave much inferior results), the penalty λ and the importance coefficient κ within the range of 0 and 1, with a step of 0.01. At the end of this process the optimum parameters that give the best classification result (based on AUC) are estimated, which in turn give the final feature subset for the respective fold. In regards to boruta, for each fold, the maximum number of importance source runs was set to 10000 and the confidence level of p-value to 0.01.

The **validation process** starts after having created the 30 models in total from the elastic-net, the regularised random forests, and boruta, for each combination, apart from combination four. At

this stage, the final step is to find out, for each combination, which model performs best. For this a 1000 repeated 10-fold cross validation is employed, which gives an average AUC with bootstrapped confidence intervals and a ROC curve with box plots in different points on the curve (60 classification models for each combination 1-3, 30 for logistic regression and 30 for random forests). The best five models of each combination are presented in the results section.

In regards to **combination 4**, the aim is to figure out whether a classification system - for diagnosing the disease and the progression - that combines the three main scenarios that are of our interest, is plausible. These scenarios refer to a) the non-progressors's group, b) the progressors' group (years of diabetes-NoDR) and c) the group of subjects with DR. This combination is important, because the progressors' group represents the pre-DR situation (we know from our retrospective data the exact periods) which we would ideally like to know, but in real-world situations we actually do not. Given that DR is just round the corner, possible treatment interventions to slow down the proliferation might be feasible. Therefore, such a diagnostic classification system, could be used to identify not only the transition from the early stage of diabetes-NoDR to pre-DR (or alternatively progressors' state), but also to DR.

The one versus all method [66] reduces the problem from multi-class to multiple binary classification problems, and was preferred due to its reduced complexity and the few available observations that are not enough for building a meaningful multi-class classifier. The process starts by training a single classifier per class, with the samples of this class as positive and the rest as negative. To achieve this, for each combination of classifiers (in this case three), the probabilities must be returned instead of the predicted class labels. The final prediction is then performed by executing these binary classifiers and finally choosing the prediction instance with the highest confidence score (probability). Similar to the other combinations, the evaluation of this ensemble of individual models is fulfilled by a 1000 repeated 10 fold cross-validation, getting one ROC curve for each of the three cases. Three feature subsets were used to test this ensemble, which are the best feature subsets of the combinations 1, 2 and 3.

3. Results and Discussion

The results that follow shed light on some not previously investigated situations, such as the progression from diabetes-noDR to DR within a progressors group across a period of time, and the simultaneous evaluation of a vast collection of geometric features for their discriminative potential. Following the methodology and approaches that were described in the previous sections, the results

of each of the six categories of Table 1, as well as those of the combinations of the classifications, are included in this section and in the supplementary materials.

3.1 Statistical Analysis

All the tables that are quoted below can be found in appendix A of the supplementary materials.

3.1.1 Four Year Matched Study of the Progressors

This group represents the main study of this work, which all the other categories are compared with. The main hypothesis suggests that the alterations in the retinal vascular geometry during the last stages of the eye with diabetes-NoDR are distinct; but given the diversity of the retina, a proper design and analysis are needed, in order to make robust and solid inferences. In this study, the same subjects and segments are used across the four year period. Therefore the whole study is absolutely balanced. The purpose is to show a) how different the results can be depending on the design of an experiment that aims to study the effects of the disease in the retinal vasculature and b) make the statistical inferences about the progression to DR.

Table A.1 shows the results of the main features, followed by Tables A.2 and A.3, which include the results of the tortuosity metrics and the post-hoc analysis, respectively.

3.1.2 Four Year Non-Matched Study of the Progressors

This part includes the same data as study 1, but the segments are not registered, so all the measurements in the main features are used without any selection. The purpose is to find out whether any differences exist in the results in comparison with the similar study 1, to justify or not the use of registered segments. Tables A.4 and A.5 show the results of the main features and the post-hoc analysis, accordingly.

Comparing Tables A.1 and A.4, it is apparent as to why not matching the segments can lead to controversial results. A series of important features that define the vascular changes, such as arterial and venular widths, were missed in the non-matched (non-registered bifurcations) design. The same exact segments, compared over a period of time, can provide more accurate representation of the underlying changes, regardless of the direction of the effect. A non significant result can be turned into a significant one with the presence of a few outliers in the sample.

3.1.3 Four Year Matched Study of the Non-Progressors

This study investigates a group of non-progressors, i.e. subjects with diabetes-NoDR that at the moment of the data collection had not progressed to DR. They are followed over a consecutive four year period of time, and the aim is to establish, whether in this period, any significant alterations due to diabetes occur. Table A.6 shows the results of the main features, followed by Tables A.7 and A.8, which include the results of the tortuosity metrics and the post-hoc analysis, respectively.

3.1.4 Four Year Non-Matched Study of the Non-Progressors

Similar to study 2, this particular study is the non-matched (non-registered bifurcations) study of study 3, using the same data but without any matching of the segments whatsoever. Accordingly, the purpose is to investigate whether any of the differences in study 3 are also observed here, and vice versa. However, no significant results were found, suggesting that, similarly to study 2, using the same bifurcations and comparing them over a period of time, vascular alterations can be identified that could have been missed otherwise (Table A.9).

3.1.5 Progressors Versus Non-Progressors Versus DR

This particular study is very important, because it averages all the years within the progressors and non-progressors groups, creating one group for each of them. Since the registered segments are used in both of them, the final values are the average of the same segments across the relevant period of time. For instance, in progressors group, the three years of measurements are averaged segment-wise. Finally, in the two groups that are created, a third one, also independent group, is included, which represents a DR group. Therefore, the analysis is run including three independent groups. The results of this novel experiment will show us two things. First, whether either of the progressors or non-progressors groups statistically differ compared to the DR group, and second whether they also differ between each other. Given that when selecting a cohort with current data it is impossible to have any information about the future progression, it is likely that many of the selected subjects will soon progress to DR, i.e. belonging to a progressors group (late stage of diabetes-NoDR). Therefore, with this study the aim is to investigate, whether a random group of subjects with diabetes but no DR (non-progressors) differs from a carefully selected progressors group, and therefore intonate the importance of more carefully selecting the cohorts.

Table A.10 shows the results of the main features, followed by Tables A.11 and A.12, which include the results of the tortuosity measurements and the post-hoc analysis, accordingly. It comes as no

surprise that the vascular geometry, both in bifurcation- and image- level, shows differences across this three group study. It also confirms the initial hypothesis, that the functional impairment and the vascular adjustments do not only occur when approaching DR, but also within the period of diabetes-NoDR. It is clear that the changes start much earlier, coming up with some strong and solid indications that diabetes starts affecting the retinal vessels quite some time before the first lesions appear.

3.1.6 Progressors Four Year Study - Independent Groups

This section is focusing on presenting the results in a similar way to study 1, with just a difference on the design. The subjects in each group are not the same across all groups. Nonetheless, it is known that they are coming from the same cohort as in study 1, so they can be compared directly with each other and make inferences about whether the results are affected, when the subjects are different. Due to the limited amount of data, the analysis is restricted to the main features. In case the matching of the segments (i.e. the proposed and suggested way of analysis - study 1) does not make any difference, then we should be able to observe the same or quite similar results with this type of design. However, as can be seen in Table A.13, this is not the case, as no statistically significant results were observed. However, this study shows something that should be stressed further. It is easy to conclude and make decisions based on misleading results. The same features, with the same dataset, but slightly differently analysed, change the whole interpretation and direction of the results. Needless to mention that the retina can drive us to more easily committing these errors, bringing again to the surface the importance of carefully selecting representative amount of large and small vessels, if matching the same segments is not an option.

In addition, as can be seen in study 5, these results indicate something extremely important. Two groups, that both include subjects with diabetes-NoDR and representative amount of measurements, are statistically significantly different across many features, such as CRVE, CRAE, venular and arterial widths. This suggests something that has to be taken into account in future studies. The selection of a cohort that is affected by diabetes has to have stricter requirements, since the alterations in the retinal vasculature are shown to change during the progression of diabetes, in particular near the onset of DR.

The ICC values reported in the tables in appendix A show whether the random effect is actually present in our data, justifying the use of LMMs. Moreover, the Ω^2 can give us an additional magnitude of the explained variance and the effect of the random covariate and the additional grouping. Most of the significant features in the above categories are observed to have fair values of Ω^2 .

To sum up, the statistical analyses of this novel study provide some very important findings. Two different cohorts (analysis of study 5) can give totally different results, which can lead to contradicting findings. This enhances the importance of studying the same subjects, and as far as possible, the same exact segments/junctions over a period of time. Planning an experiment, and recruiting subjects with diabetes, one cannot know at the moment of the recruitment whether some of these subjects will progress to DR very soon after. This suggests that the variability within the cohort might be high enough to be possible to identify subtle vascular changes, when compared to a DR group. This difference could also suggest one additional thing. The cohort of subjects with diabetes-NoDR (progressors group) that is about to progress to DR in the near future, presents some noticeable differences, compared to another non-progressors group, which actually supports the hypothesis that the retinal vasculature is affected and adapts to the underlying functional alterations that occur right before the onset of DR. Examples of that can be seen in Tables A.10 and A.11, where many of these features, such as CRAE, arterial widths and tortuosity, differ significantly between the two cohorts. The findings can partly be explained by the fact that microaneurysms are formed due to the loss of pericytes, which apparently precedes the appearance of lesions, whereas vascular occlusion is a consequence of basement membrane thickening and endothelial cell proliferation [67]. These processes, which are initially subtle, affect the retinal haemodynamics. In the normal retina, blood flow remains constant by the autoregulated vascular responses over a range of systemic blood pressures and intra-ocular pressures [68]. Local factors, which target smooth muscle cells in arterioles and capillary pericytes in subjects with diabetes are controlling the vessels [69]. In the progressors group, prior to the onset of DR, the significantly altered features, e.g. CRVE, CRAE, fractal dimension, venular and arterial widths, are following a mixed trend in terms of increase/decrease of mean values. On the other hand, the generalised tortuosity is increased, suggesting that the progression from diabetes to DR is linked to more tortuous vessels.

3.2 *Classification Models*

Following the statistical evaluation of the features, which gives us a specific type of information, this section aims to utilise these features in practice and also investigate whether a subset of them can be useful in discriminating between groups of different cohorts. For each combinations 1-3 the results of the feature selection process and the classification models are presented. For each combination, the best five models and feature subsets are presented.

3.2.1 Combination 1

This combination includes two classes; one is the DR group and the other is the average of the entire three year period pre-DR. This combination attempts to evaluate the changes that have occurred to the retinal vasculature between the last stages of the eye with diabetes-NoDR and the onset of DR. The best performance in this combination was achieved by the feature subset and classification as delivered by the regularised random forest's classifier (Figure 8, mtry=4, $\lambda=0.89$), yielding an AUC of 0.878 (Table 5, Figure 9), and positive predictive value (PPV) and negative predictive value (NPV) of 88.34% and 85.26% respectively.

Table 5: Summary of the classification performance for Combination 1

Summary	Best model	Model 2	Model 3	Model 4	Model 5
Random Forests	0.878	0.865	0.813	0.647	0.672
AUC (CI95%)	(0.753-0.954)	(0.764-0.958)	(0.654-0.921)	(0.489-0.777)	(0.543-0.798)
Logistic Regression	0.735	0.675	0.754	0.763	0.724
AUC (CI95%)	(0.534-0.853)	(0.521-0.765)	(0.567-0.843)	(0.574-0.853)	(0.531-0.834)
Random Forests	0.572	0.524	0.511	0.245	0.269
kappa					
Logistic Regression	0.468	0.343	0.503	0.509	0.462
kappa					
Random forests	0.203	0.234	0.245	0.367	0.389
OOB					
Feature selection process	RF	BO	RF	EN	EN

The features of the best model for combination 1, ordered according to their importance, are the following:

Best model: Angle.BC_Art, VeinBc, mean_psi, Artgamma, CRAE_Leon, VeinAnglSD, sd_phi, veingamma, X75th_phi, median_tau.

Model 2: VeinBc, Angle.BC_Art, mean_psi, Artgamma, CRAE_Leon, VeinAnglSD, Fractal.

Model 3: Fractal, Artgamma, CRAE_Leon, VeinAnglSD.

Model 4: Fractal, Angle.BC_Vein, Angle.BC_Art, CRAE_Leon.

Model 5: Fractal, Angle.BC_Vein, Angle.BC_Art, MedAngArt, ArtAngleSD.

The performance suggests that still distinct changes are occurring that could differentiate between late stage diabetes-NoDR and DR cohorts, relying only on the vascular geometry (an example can be seen in Figure 10). In addition, the good performance of the classifier implies that it could be used for classifying a progressors cohort versus a DR one, since the three year period of available information is adequate to create a representative sample.

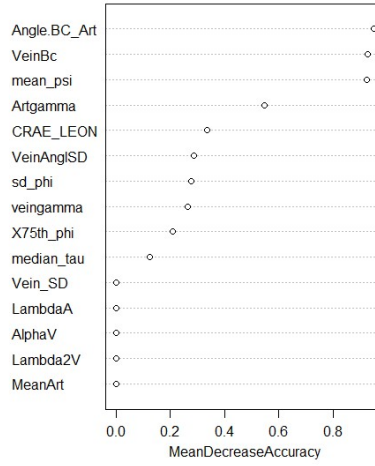


Figure 8: Ranking of the features of the best model for combination 1, as returned from the random forests feature selection process.

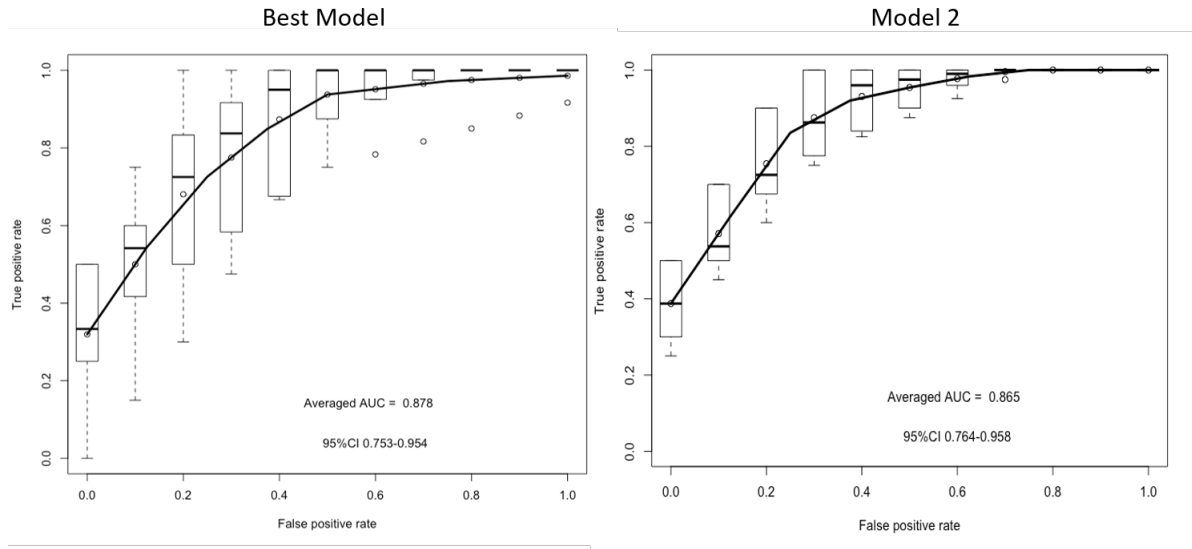


Figure 9: On the left: ROC curve and AUC of the best model with the confidence intervals for combination 1; on the right similar information is given for the second best model. This plot is the averaged ROC curve over all the cross-validation runs, including the variation and the spread estimates around the average curve using box plots. The spread estimates are shown at 11 equally spaced positions along the curve.

3.2.2 Combination 2

In this combination, the purpose is to evaluate how a classifier can possibly perform in a realistic situation. If a high performance, reliable and clinically-validated classification model is to be proposed that could differentiate between subjects with diabetes-NoDR and DR, it needs to include independent

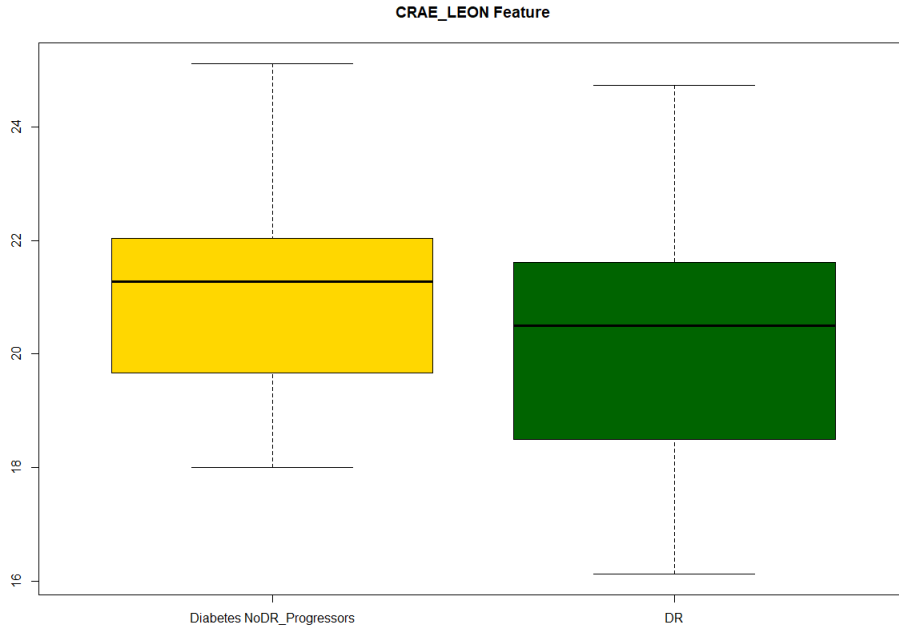


Figure 10: An example boxplot showing the CRAE_LEON feature in both classes for combination 1.

information, given that the future instances that will have to be classified will come from independent subjects. To test this scenario, the classifier was trained by including information from the non-progressors and DR groups.

The maximum achieved performance was delivered by a feature subset of boruta and the logistic regression classifier, giving an AUC of 0.968 (Table 6, Figures 11 and 12), PPV of 96.78% and NPV 98.35%.

Table 6: Summary of the classification performance for Combination 2

Summary	Best model	Model 2	Model 3	Model 4	Model 5
Random Forests	0.945	0.961	0.943	0.942	0.931
AUC (CI95%)	(0.817-0.991)	(0.869-0.992)	(0.835-0.987)	(0.844-0.975)	(0.812-0.972)
Logistic Regression	0.968	0.823	0.786	0.794	0.903
AUC (CI95%)	(0.923-0.999)	(0.656-0.912)	(0.643-0.889)	(0.587-0.887)	(0.789-0.981)
Random Forests	0.887	0.907	0.875	0.867	0.854
kappa					
Logistic Regression	0.891	0.654	0.567	0.587	0.803
kappa					
Random forests	0.113	0.124	0.133	0.122	0.127
OOB					
Feature selection process	BO	BO	BO	RF	EN

The features of the best model for combination 2, ordered according to their importance, are the following:

Best model: CRAE_KNUD, CRAE_LEON, CRVE_LEON, arterySD, Artgamma, CRVE_KNUD, BC_Mean_Art, ArtAngleSD, VeinBc, sd_phi, MedAngVein, MedArt, VeinAnglSD, veingamma, MedVein, Angle.BC_Art, X75th_phi, Lambda1V, Tort, mean_phi

Model 2: CRAE_LEON, CRAE_KNUD, CRVE_KNUD, arterySD, Artgamma, VeinAnglSD, CRVE_LEON, Medart, VeinBc, ArtAnglSD, BC_Mean_Art.

Model 3: CRAE_LEON, CRAE_KNUD, CRVE_LEON, Artgamma, arterySD, sd_phi, BC_Mean_Art, MedAngVein, ArtAngleSD, Medvein, MedArt, CRVE_KNUD, veingamma.

Model 4: CRAE_LEON, MedAngVein, arterySD.

Model 5: MedVein, VeinAngSd, CRAE_KNUD, CRAE_LEON, CRVE_KNUD, CRVE_LEON, sd_phi, Tort.

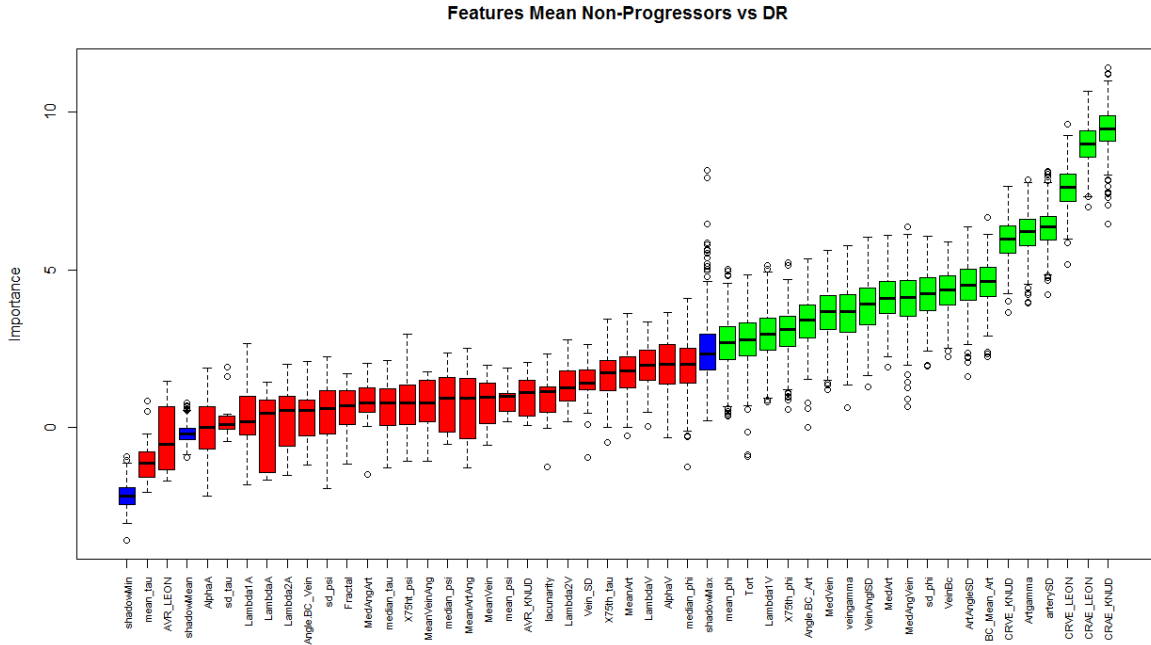


Figure 11: Feature selection process with boruta for the best model of combination 2; y axis shows the relative importance of each feature, whereas x axis includes the features. Green coloured features are the finally selected ones, whereas the red are the rejected ones.

The interesting part of this combination is that these features are different from the ones in combination 1. This suggests that there are different attributes that differentiate the two diverse cohorts (non-progressors and progressors) from DR, implying that the vascular changes are not the same across different periods of time as the disease proliferates. This is something that is also supported by the statistical analysis. Therefore it is worth investing on collecting representative data from both these cohorts, and also make all the efforts to be considered separately in prospective studies. The remarkable

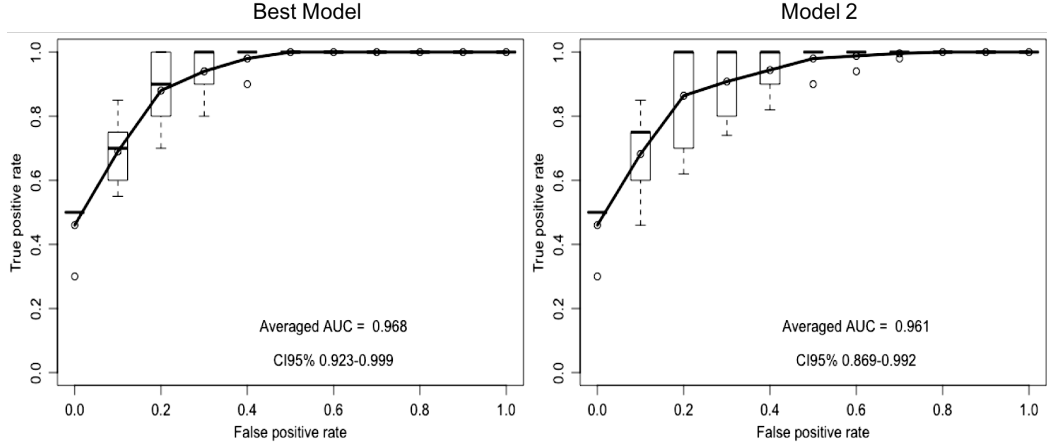


Figure 12: On the left: ROC curve and AUC of the best model with the confidence intervals for combination 2; on the right similar information is given for the second best model.

performance of this classification model, with the tight confidence intervals as well, provides evidence that such a model could be used for differentiating between subjects with diabetes-NoDR and subjects with DR.

3.2.3 Combination 3

Similar to the objective of the statistical analysis of study 2, this combination aims to efficiently classify a group of subjects with diabetes that are non-progressors, and a group that they are about to progress to DR soon. This combination has high clinical importance, given that being able to accurately differentiate between these two groups, can give important indications that one is about to progress to DR in the near future. The random forests classifier gave the best performance of 0.963 in AUC (Table 7, Figure 14), 98.77% in PPV and 97.15% NPV, using a feature subset that was selected by the boruta feature selection process (Figure 13).

Table 7: Summary of the classification performance for Combination 3

Summary	Best model	Model 2	Model 3	Model 4	Model 5
Random Forests	0.963	0.957	0.955	0.949	0.937
AUC (CI95%)	(0.898-0.998)	(0.887-0.991)	(0.893-0.997)	(0.896-0.987)	(0.867-0.977)
Logistic Regression	0.923	0.865	0.902	0.845	0.856
AUC (CI95%)	(0.854-0.995)	(0.667-0.934)	(0.756-0.983)	(0.673-0.921)	(0.667-0.946)
Random Forests	0.903	0.896	0.856	0.843	0.823
kappa					
Logistic Regression	0.834	0.723	0.805	0.709	0.715
kappa					
Random forests	0.092	0.098	0.107	0.111	0.126
OOB					
Feature selection process	BO	RF	RF	RF	RF

The features of the best model for combination 3, ordered according to their importance, are the following:

Best model: CRVE_LEON, CRAE_LEON, ArtAngleSD, CRVE_KNUD, arterySD, CRAE_KNUD, BC_Mean_Art, Artgamma, MedArt, VeinAnglSD, MeanArt, Angle_BC_Art, sd_phi, MedVein, Angle_BC_Vein, MedAngVein.

Model 2: CRVE_LEON, CRVE_KNUD, CRAE_LEON, ArtAngSD, Veingamma, Medvein.

Model 3: CRVE_LEON, ArtAngSD, MeanVeinAng, mean_tau.

Model 4: CRVE_LEON, BC_Mean_Art, CRAE_LEON, ArtAngleSD, VeinAnglSD, MeanArtAng, sd_phi.

Model 5: CRAE_LEON, artAngSD, CRVE_LEON, Angle_BC_Art, CRVE_KNUD.

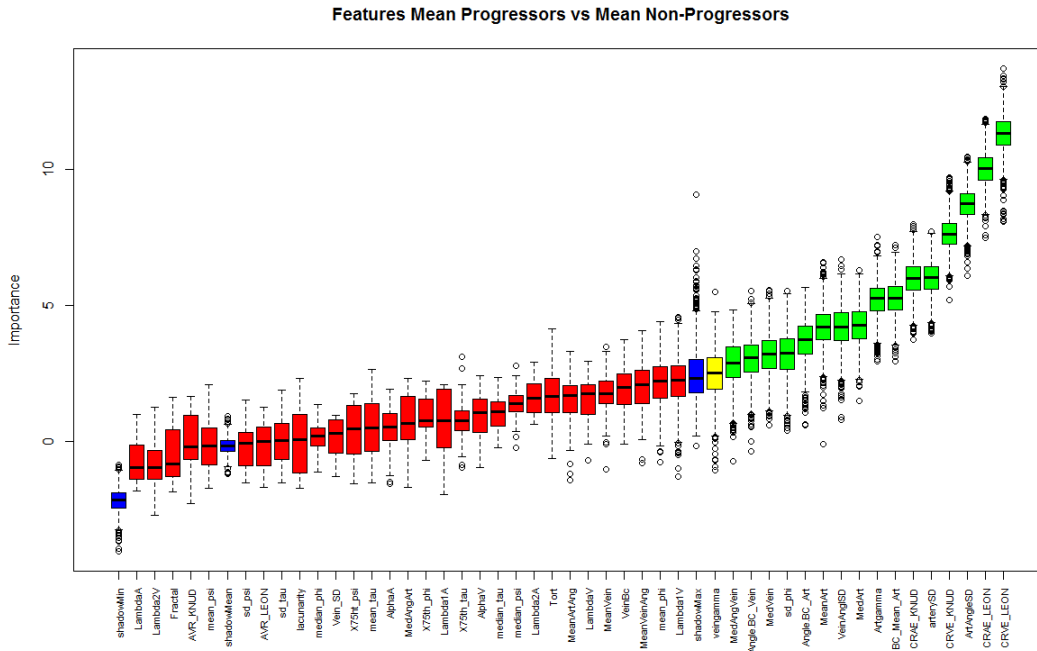


Figure 13: Feature selection process with boruta for the best model of combination 3; y axis shows the relative importance of each feature, whereas x axis includes the features. Green coloured features are the finally selected ones, whereas the red are the rejected ones. In yellow colour is a tentative feature, which is not significantly important (with respect to a p -value ≤ 0.05).

Just as has been indicated by the statistical analysis, it comes as no surprise that features such as CRAE, CRVE, tortuosity, arterial and venular widths and angles yield a very good performance in detecting the stages of the disease. This suggests that such a classifier could be used in the near future for identifying the progression of a-not-yet-progressed to DR subject with diabetes. This is an important observation, which could be paramount for managing to have a clinical system in the future

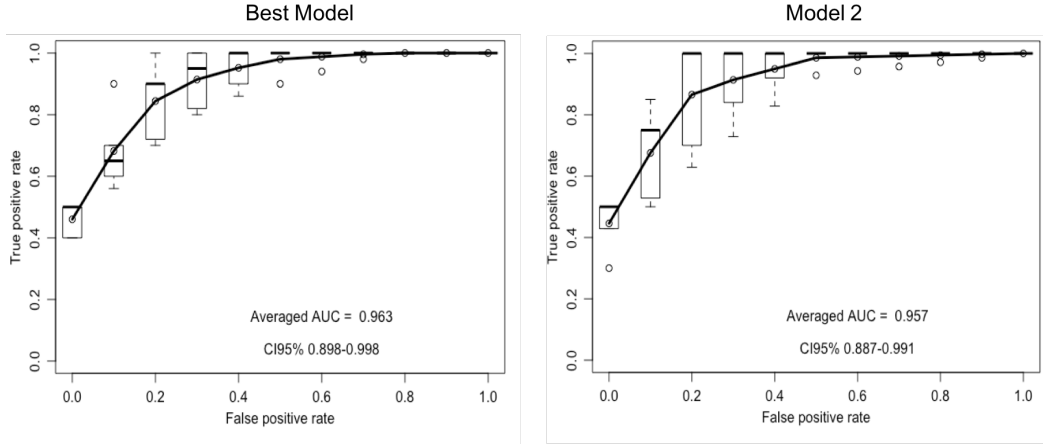


Figure 14: On the left: ROC curve and AUC of the best model with the confidence intervals for combination 3; on the right similar information is given for the second best model.

that can assist with the assessment of the disease’s progression, prior to the appearance of lesions. It goes without saying that a clinical exploitation is contingent upon a much larger sample of data than what was available in these experiments, in order to investigate the possibility of deploying such a system in a clinical environment.

3.2.4 Combination 4 - Pilot Study

This final combination presents a pilot classification model, in order to investigate whether a system that can successfully classify the three crucial stages of the early progression could be feasible. Such a model can have a direct application in identifying the stage of the disease of a subject, and/or the probability that he/she belongs to a specific stage. Just like in combinations 2 and 3, it is paramount to be able to detect the stage of the disease, and possibly assist the clinicians in the decision making process.

As previously mentioned, in this specific comparison, the best feature subsets from the combinations 1, 2, 3 were used to train the random forests classifier. The best result was obtained by the feature subset of combination 3 (AUC = 0.821). Figure 15 shows the ROC curves of all the combinations, as well as the individual ROC curves for each binary classifier. The performance is decent enough to suggest that it is possible to identify these important stages of the disease. It is also promising that such a performance was achieved with limited observations and only with information of the retinal geometry. The further evaluation of such a classifier in a clinical study is definitely useful.

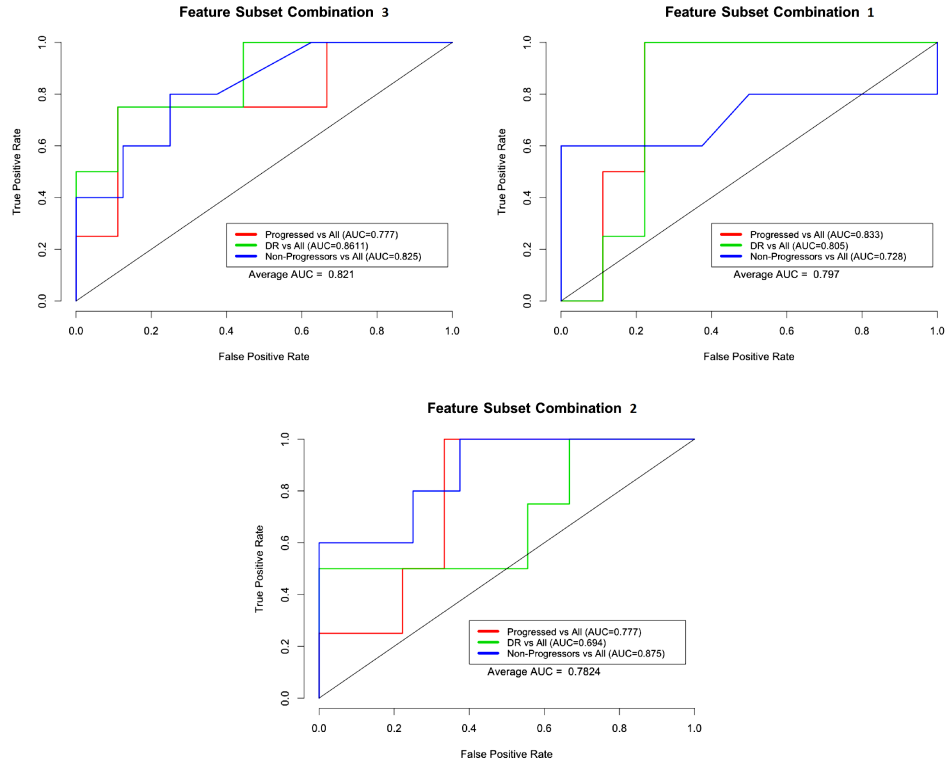


Figure 15: ROC curves for each of the three combinations used in building the one vs all classification models. For each feature subset, the plot includes three ROC curves that correspond to the three different possible classifications: progressors versus the rest, non-progressors versus the rest and DR versus the rest. The upper left plot of combination 3 presents the best results, followed by the one on the upper right of combination 1, and finishing with the one down of combination 2.

3.2.4.1 Feature Ranking

The top 14 features as they appeared in all of the feature selection processes (90 in total), across all the different combinations, can be seen in Table 8. This table provides us with very important

Table 8: Ranking of the 14 most selected features

Features	Appearances in Models (out of 90)	Appearances in Bootstrap
ArtAngleSD	39	45.32%
CRAE_LEON	36	44.14%
CRVE_LEON	35	44.22%
Angle_BC_art	32	40.92%
CRVE_KNUD	31	38.45%
ArterySD	29	36.88%
CRAE_KNUD	27	33.1%
VeinAnglSD	27	32.65%
BC_Mean_Art	24	30.36%
Artgamma	20	25.79%
Fractal	20	26.56%
Angle_BC_Vein	19	25.31%
MedVein	17	21.4%
Sd_phi	17	20.99%

information, showing that the retinal vascular geometry, and mainly the vessel widths, vessel angles and tortuosity are strong features, with good discriminative potential, across all the combinations that were investigated. As was observed both in the statistical analysis and the classification models, the features of Table 8 shall be taken into account when building a detection system.

3.3 Discussion

Undoubtedly, the retinal vascular geometry is shown to be changing during the progression of diabetes, which can possibly be attributed to processes that include, among others, the neurodegeneration, the subclinical inflammation, the oxidative stress and the dysfunction of the endothelial cells that hyperglycaemia also causes [7]. In addition, the dysfunction of the endothelial cells affects the production of nitric oxide in response to shear stress, which can lead to the development of atherosclerosis, i.e. the hardening and thickening of the arterial wall, thus the observed arterial alterations.

As it was hypothesised, matching the vessel segments and studying them over a period of time, one can identify changes that might be missed otherwise. All of the above results strongly suggest that it is very easy to be misled and come up with conclusions that are not representative of what is being studied. Nonetheless, the experiments showed that the vascular geometry is indeed affected during the progression of diabetes, which is more or less in line with some of the previous findings [2, 17, 70]. The vessel widths, FD, venular angle to BC ratio, CRVE and CRAE were found to be important discriminative features of progression to DR. CRVE and CRAE observations are in line with the findings of Klein [71] and partly with [14, 72, 73]. These observations are linked possibly to the endothelial dysfunction, inflammatory changes, and hyperglycemia, all of which are factors involved in the pathogenesis of DR [74, 75]. As can be seen in Table A.10, not only is there an increase to both these features (CRVE and CRAE) between the groups of non-progressors and DR, but also the same occurs between the non-progressors and progressors. This extends the observations of Klein *et al.* [71], suggesting that the changes are not necessarily occurring during or around the DR period, but also few years earlier than the onset of DR, when the retina is adjusting to the upcoming vascular and functional alterations. Fractal dimension is also an important feature, with significant changes between the two cohorts (progressors/non-progressors) and DR, which comes in contrast to the findings in [15].

The results of the classification models suggest that detection systems built solely with information of the retinal geometry can possibly provide reliable systems that could be tested and incorporated in a real-world clinical application. These features accommodate the purpose of this work, which is the early detection of the progression from diabetes-NoDR to DR. Therefore, these novel findings have a

clear clinical interest.

It needs to be kept in mind that the features that appear in the feature selection processes are those that can provide a better classification performance, as defined by the corresponding techniques. This means that highly correlated features that have similar performance might not be selected. Nonetheless, Table 8 is a strong indicator of which features have good discriminative potential that should be considered when designing diagnostic systems.

Another important observation is that extensive changes between the non-progressors and the progressors groups exist, such as in the CRVE, CRAE, the venular widths and tortuosity. This is one of the key findings of this work, suggesting that changes occur early, much before DR, and continue until the years before the onset of DR, adopting at this stage the characteristics of a progressors' group (instead of a non-progressors). Following that, more alterations are under way, until the first lesions finally appear in the retina. The findings for the CRAE and CRVE are in line with the ones in [71, 73].

In regards to the classification models, the key and rigorous outcome of this novel study is that it was succinctly shown that a reliable and high performance classifier can indeed be used to differentiate between different stages of the disease's progression, implemented only with vascular geometric features.

4. Conclusion and Future Work

A novel machine learning-based framework and a comprehensive study were presented that combine a number of geometric features from the retinal vasculature, for detecting the progression to DR. It was explicitly shown that many features have good discriminative potential. Subject to proper clinical validation, a decision support system is a feasible target that can assist with the diagnosis of the progression to DR. The performance of the different classification systems, as presented in this work, clearly support this statement.

As for the future steps, the author acknowledges the importance of collecting additional data that can attest and corroborate the findings. Working closely with clinical partners, the efforts are placed on obtaining such data and apply the FRiAReD on a large scale study, including its clinical validation. Second, we aim to work on fully automating the framework, in particular the artery/vein classification.

Acknowledgments

This research project was partly supported by a Marie Skłodowska-Curie grant from the European Commission in the framework of the REVAMMAD ITN (Initial Training Research network), Project number 316990. I would like to thank Professor Andrew Hunter and Dr Bashir Al-Diri for their advice on some aspects of the project.

Conflict of Interest statement

Author has no conflict of interest to declare.

Appendix A

See Supplementary Materials.

References

- [1] M. S. Haleem, L. Han, J. Van Hemert, A. Fleming, L. R. Pasquale, P. S. Silva, B. J. Song, L. P. Aiello, Regional image features model for automatic classification between normal and glaucoma in fundus and scanning laser ophthalmoscopy (slo) images, *Journal of medical systems* 40 (6) (2016) 1–19.
- [2] G. Leontidis, B. Al-Diri, J. Wigdahl, A. Hunter, Evaluation of geometric features as biomarkers of diabetic retinopathy for characterizing the retinal vascular changes during the progression of diabetes, in: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 5255–5259. doi:10.1109/EMBC.2015.7319577.
- [3] J. Wigdahl, P. Guimarães, G. Leontidis, A. Triantafyllou, A. Ruggeri, Automatic gunn and salus sign quantification in retinal images, in: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, IEEE, 2015, pp. 5251–5254.
- [4] S. B. Seidelmann, B. Claggett, P. E. Bravo, A. Gupta, H. Farhad, B. E. Klein, R. Klein, M. F. Di Carli, S. D. Solomon, Retinal vessel calibers in predicting long-term cardiovascular outcomes: The atherosclerosis risk in communities study, *Circulation* (2016) CIRCULATIONAHA–116.
- [5] S. S. Hayreh, B. Zimmerman, M. J. McCarthy, P. Podhajsky, Systemic diseases associated with various types of retinal vein occlusion, *American journal of ophthalmology* 131 (1) (2001) 61–77.

- [6] W. H. Organization, Global report on diabetes, 2016.
- [7] G. Leontidis, B. Al-Diri, A. Hunter, Diabetic retinopathy: current and future methods for early screening from a retinal hemodynamic and geometric approach, *Expert Review of Ophthalmology* 9 (5) (2014) 431–442.
- [8] R. Donnelly, E. Horton, Vascular complications of diabetes: current issues in pathogenesis and treatment, John Wiley & Sons, 2008.
- [9] G. Leontidis, Early screening and diagnosis of diabetic retinopathy, Ph.D. thesis, University of Lincoln (2016).
- [10] W. S. Oliveira, J. V. Teixeira, T. I. Ren, G. D. Cavalcanti, J. Sijbers, Unsupervised retinal vessel segmentation using combined filters, *PloS one* 11 (2) (2016) e0149943.
- [11] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, T. Wang, A cross-modality learning approach for vessel segmentation in retinal images, *IEEE transactions on medical imaging* 35 (1) (2016) 109–118.
- [12] A. Perez-Rovira, T. MacGillivray, E. Trucco, K. Chin, K. Zutis, C. Lupascu, D. Tegolo, A. Giachetti, P. Wilson, A. Doney, et al., Vampire: vessel assessment and measurement platform for images of the retina, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 3391–3394.
- [13] E. Grisan, M. Foracchia, A. Ruggeri, A novel method for the automatic grading of retinal vessel tortuosity, *IEEE transactions on medical imaging* 27 (3) (2008) 310–319.
- [14] X. Yang, Y. Deng, H. Gu, X. Ren, A. Lim, T. Snellingsen, X. Liu, N. Wang, J. W. Pak, N. Liu, et al., Relationship of retinal vascular calibre and diabetic retinopathy in chinese patients with type 2 diabetes mellitus: the desheng diabetic eye study, *British Journal of Ophthalmology* (2016) bjophthalmol–2014.
- [15] A. Kunicki, A. Oliveira, M. Mendonça, C. Barbosa, R. Nogueira, Can the fractal dimension be applied for the early diagnosis of non-proliferative diabetic retinopathy?, *Brazilian Journal of Medical and Biological Research* 42 (10) (2009) 930–934.
- [16] C. Y.-l. Cheung, E. Lamoureux, M. K. Ikram, M. B. Sasongko, J. Ding, Y. Zheng, P. Mitchell, J. J. Wang, T. Y. Wong, Retinal vascular geometry in asian persons with diabetes and retinopathy, *Journal of diabetes science and technology* 6 (3) (2012) 595–605.

- [17] G. Leontidis, B. Al-Diri, A. Hunter, Exploiting the retinal vascular geometry in identifying the progression to diabetic retinopathy using penalized logistic regression and random forests, in: *Emerging Trends and Advanced Technologies for Computational Intelligence*, Springer, 2016, pp. 381–400.
- [18] Z. Burgansky-Eliash, A. Barak, H. Barash, D. A. Nelson, O. Pupko, A. Lowenstein, A. Grinvald, A. Rubinstein, Increased retinal blood flow velocity in patients with early diabetes mellitus, *Retina* 32 (1) (2012) 112–119.
- [19] F. Caliva, G. Leontidis, P. Chudzik, A. Hunter, L. Antiga, B. Al-Diri, et al., Hemodynamics in the retinal vasculature during the progression of diabetic retinopathy, *Journal for Modeling in Ophthalmology* 1 (4) (2017) 6–15.
- [20] S. Roychowdhury, D. D. Koozekanani, K. K. Parhi, Dream: Diabetic retinopathy analysis using machine learning, *IEEE journal of biomedical and health informatics* 18 (5) (2014) 1717–1728.
- [21] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, M. Niemeijer, Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning, *Investigative Ophthalmology & Visual Science* 57 (13) (2016) 5200–5206.
- [22] M. D. Abràmoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, P. Massin, B. Cochener, P. Gain, L. Tang, et al., Automated analysis of retinal images for detection of referable diabetic retinopathy, *JAMA ophthalmology* 131 (3) (2013) 351–357.
- [23] C. Wilkinson, F. L. Ferris, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, J. T. Verdager, et al., Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales, *Ophthalmology* 110 (9) (2003) 1677–1682.
- [24] Matlab, version 8.4.0 (R2014b), The MathWorks Inc., Natick, Massachusetts, 2014.
- [25] RStudio Team, RStudio: Integrated Development Environment for R, RStudio, Inc., Boston, MA (2015).
URL <http://www.rstudio.com/>
- [26] R. C. of Ophthalmologists, Diabetic retinopathy guidelines (2012).

- [27] A. Vedaldi, B. Fulkerson, VLFeat: An open and portable library of computer vision algorithms, <http://www.vlfeat.org/> (2008).
- [28] D. G. Lowe, Object recognition from local scale-invariant features, in: Computer vision, 1999. The proceedings of the seventh IEEE international conference on, Vol. 2, Ieee, 1999, pp. 1150–1157.
- [29] D. Kriegman, Homography estimation, Lecture Computer Vision I, CSE A 252.
- [30] E. Dubrofsky, Homography estimation, Diplomová práce. Vancouver: Univerzita Britské Kolumbie.
- [31] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (6) (1981) 381–395.
- [32] B. Al-Diri, A. Hunter, D. Steel, An active contour model for segmenting and measuring retinal vessels, IEEE Transactions on Medical imaging 28 (9) (2009) 1488–1497.
- [33] B. Al-Diri, A. Hunter, D. Steel, M. Habib, Manual measurement of retinal bifurcation features, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 4760–4764.
- [34] A. W. Stitt, T. M. Curtis, Advanced glycation and retinal pathology during diabetes, Pharmacological Reports 57 (2005) 156.
- [35] A. W. Stitt, N. Lois, R. J. Medina, P. Adamson, T. M. Curtis, Advances in our understanding of diabetic retinopathy, Clinical science 125 (1) (2013) 1–17.
- [36] V. P. Singh, A. Bali, N. Singh, A. S. Jaggi, Advanced glycation end products and diabetic complications, The Korean Journal of Physiology & Pharmacology 18 (1) (2014) 1–14.
- [37] G. Leontidis, B. Al-Diri, A. Hunter, Summarising the retinal vascular calibres in healthy, diabetic and diabetic retinopathy eyes, Computers in biology and medicine 72 (2016) 65–74.
- [38] M. D. Knudtson, K. E. Lee, L. D. Hubbard, T. Y. Wong, R. Klein, B. E. Klein, Revised formulas for summarizing retinal vessel diameters, Current eye research 27 (3) (2003) 143–149.
- [39] K. Chandrinos, M. Pilu, R. Fisher, P. Trahanias, Image processing techniques for the quantification of atherosclerotic changes, Department of Artificial Intelligence, University of Edinburgh, 1998.

- [40] W. E. Hart, M. Goldbaum, B. Côté, P. Kube, M. R. Nelson, Measurement and classification of retinal vascular tortuosity, *International journal of medical informatics* 53 (2) (1999) 239–252.
- [41] E. Poletti, E. Grisan, A. Ruggeri, Image-level tortuosity estimation in wide-field retinal images from infants with retinopathy of prematurity., in: *Conference proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, Vol. 2012, 2011, pp. 4958–4961.
- [42] M. Zamir, Nonsymmetrical bifurcations in arterial branching., *The Journal of general physiology* 72 (6) (1978) 837–845.
- [43] R. Karch, F. Neumann, M. Neumann, W. Schreiner, Staged growth of optimized arterial model trees, *Annals of biomedical engineering* 28 (5) (2000) 495–511.
- [44] M. A. Mainster, The fractal properties of retinal vessels: embryological and clinical implications, *Eye* 4 (1) (1990) 235–241.
- [45] C. R. Tolle, T. R. McJunkin, D. J. Gorsich, An efficient implementation of the gliding box lacunarity algorithm, *Physica D: Nonlinear Phenomena* 237 (3) (2008) 306–315.
- [46] B. B. Mandelbrot, *The fractal geometry of nature*, Vol. 173, Macmillan, 1983.
- [47] K. Foroutan-Pour, P. Dutilleul, D. Smith, Advances in the implementation of the box-counting method of fractal dimension estimation, *Applied mathematics and computation* 105 (2) (1999) 195–210.
- [48] T. T. Nguyen, J. J. Wang, T. Y. Wong, Retinal vascular changes in pre-diabetes and prehypertension new findings and their research and clinical implications, *Diabetes care* 30 (10) (2007) 2708–2715.
- [49] C. E. McCulloch, J. M. Neuhaus, *Generalized linear mixed models*, Wiley Online Library, 2001.
- [50] D. Bates, M. Maechler, B. Bolker, S. Walker, et al., lme4: Linear mixed-effects models using eigen and s4, *R package version 1* (7).
- [51] E.-J. Wagenmakers, A practical solution to the pervasive problems of p values, *Psychonomic bulletin & review* 14 (5) (2007) 779–804.

- [52] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, G. Grothendieck, P. Green, M. B. Bolker, Package lme4 (2016).
- [53] R. J. Feise, Do multiple outcome measures require p-value adjustment?, BMC Medical Research Methodology 2 (1) (2002) 1.
- [54] A. Gelman, J. Hill, M. Yajima, Why we (usually) don't have to worry about multiple comparisons, Journal of Research on Educational Effectiveness 5 (2) (2012) 189–211.
- [55] R. Xu, Measuring explained variation in linear mixed effects models, Statistics in medicine 22 (22) (2003) 3527–3541.
- [56] H. Deng, G. Runger, Gene selection with guided regularized random forest, Pattern Recognition 46 (12) (2013) 3483–3489.
- [57] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, Journal of statistical software 33 (1) (2010) 1.
- [58] M. B. Kursa, W. R. Rudnicki, et al., Feature selection with the boruta package (2010).
- [59] P. A. Devijver, J. Kittler, Pattern recognition: A statistical approach, Prentice hall, 1982.
- [60] B. Efron, R. Tibshirani, Improvements on cross-validation: the 632+ bootstrap method, Journal of the American Statistical Association 92 (438) (1997) 548–560.
- [61] F. J. Provost, T. Fawcett, R. Kohavi, The case against accuracy estimation for comparing induction algorithms., in: ICML, Vol. 98, 1998, pp. 445–453.
- [62] A. Ben-David, Comparison of classification accuracy using cohens weighted kappa, Expert Systems with Applications 34 (2) (2008) 825–832.
- [63] M. Kuhn, caret: Classification and Regression Training, r package version 6.0-68 (2016).
- [64] M. B. Kursa, W. R. Rudnicki, Feature selection with the Boruta package, Journal of Statistical Software 36 (11) (2010) 1–13.
- [65] T. Sing, O. Sander, N. Beerenwinkel, T. Lengauer, Rocr: visualizing classifier performance in r, Bioinformatics 21 (20) (2005) 7881.

- [66] R. Rifkin, A. Klautau, In defense of one-vs-all classification, *Journal of machine learning research* 5 (Jan) (2004) 101–141.
- [67] G. L. King, T. Shiba, J. Oliver, T. Inoguchi, S.-E. Bursell, Cellular and molecular abnormalities in the vascular endothelium of diabetes mellitus, *Annual review of medicine* 45 (1) (1994) 179–188.
- [68] C. E. Riva, S. H. Sinclair, J. E. Grunwald, Autoregulation of retinal circulation in response to decrease of perfusion pressure., *Investigative ophthalmology & visual science* 21 (1) (1981) 34–38.
- [69] G. De La Rubia, F. J. Oliver, T. Inoguchi, G. L. King, Induction of resistance to endothelin-1’s biochemical actions by elevated glucose levels in retinal pericytes, *Diabetes* 41 (12) (1992) 1533–1539.
- [70] G. Leontidis, J. Wigdahl, B. Al-Diri, A. Ruggeri, A. Hunter, Evaluating tortuosity in retinal fundus images of diabetic patients who progressed to diabetic retinopathy, in: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015.
- [71] R. Klein, B. E. Klein, S. E. Moss, T. Y. Wong, L. Hubbard, K. J. Cruickshanks, M. Palta, The relation of retinal vessel caliber to the incidence and progression of diabetic retinopathy: Xix: The wisconsin epidemiologic study of diabetic retinopathy, *Archives of ophthalmology* 122 (1) (2004) 76–83.
- [72] A. S. Tsai, T. Y. Wong, R. Lavanya, R. Zhang, H. Hamzah, E. S. Tai, C. Y. Cheung, Differential association of retinal arteriolar and venular caliber with diabetes and retinopathy, *Diabetes research and clinical practice* 94 (2) (2011) 291–298.
- [73] F. Islam, T. Nguyen, J. Wang, E. Tai, A. Shankar, S. Saw, T. Aung, S. Lim, P. Mitchell, T. Wong, Quantitative retinal vascular calibre changes in diabetes and retinopathy: the singapore malay eye study, *Eye* 23 (8) (2009) 1719–1724.
- [74] M. I. Schmidt, B. B. Duncan, A. R. Sharrett, G. Lindberg, P. J. Savage, S. Offenbacher, M. I. Azambuja, R. P. Tracy, G. Heiss, A. investigators, et al., Markers of inflammation and prediction of diabetes mellitus in adults (atherosclerosis risk in communities study): a cohort study, *The Lancet* 353 (9165) (1999) 1649–1652.

- [75] R. Klein, C. E. Myers, K. E. Lee, R. Gangnon, B. E. Klein, Changes in retinal vessel diameter and incidence and progression of diabetic retinopathy, *Archives of Ophthalmology* 130 (6) (2012) 749–755.

Appendix A. Data Extraction and Statistical Results

The data that were extracted for conducting the above six studies can be found as follows:

Study 1: Out of the 127 subjects and 508 temporal images, 852 venular junctions were successfully matched for the main features across the four year period, extracting 2556 venular width measurements. For the arteries, the total amount was 388 junctions, successfully measuring 1164 arterial widths. The image-level features of tortuosity, fractal dimension and lacunarity were estimated in the whole set of images. CRVE, CRAE and AVR features were extracted out of 200 nasal images.

Study 2: 904 arterial and 2534 venular junctions (same number for vessel angles) were included in this non-matched progressors study (2712 and 7602 arterial and venular vessel widths, respectively).

Study 3: The non-progressors' group contains 27 subjects across a four year period, including 108 temporal images. 136 arterial junctions were successfully matched, measuring 408 widths in total. Accordingly, 220 venular junctions were also matched, giving a total number of 660 width measurements. Tortuosity, fractal dimension and lacunarity were estimated in the whole set of images. CRVE, CRAE and AVR were extracted out of 80 nasal images.

Category 4: 379 arterial and 594 venular junctions were included in the non-matched non-progressors' study (1137 arterial and 1782 venular widths, respectively).

Category 5: For the arteries, 228 junctions were included, and similarly 471 venular junctions. Tortuosity, fractal dimension and lacunarity were estimated from a total 282 temporal images. CRVE, CRAE and AVR features were extracted out of 100 nasal images.

Category 6: This final study includes only the main features, but without AVR, CRVE and CRAE. The 127 subjects were randomly assigned to one of the four groups, discarding at random three of the subjects, in order to have equal numbers of subjects in each group. Therefore, each group comprises of 31 subjects, who contributed with their non-matched measurements for the analysis. The total number of junctions were 219 for arteries (657 vessel widths) and 456 for veins (1368 vessel widths). Accordingly, for the fractal dimension and lacunarity a total number of 124 images were used (31 for each of the four groups).

Appendix A.1 Four Year Matched Study of the Progressors

Table A.1: Analysis of the main features for study 1

Feature name	AIC ^a	BIC ^a	(p-(χ^2) ^b	ICC	Ω^2	Group Means(SD) (Y3, Y2, Y1, DR)
AlphaA	306.5/ 308.8	283.9/ 292.4	0.039	0.46	0.605	0.77(0.16),0.82(0.15) 0.81(0.12),0.79(0.16)
BetaA	35.4/ 36.4	20.5/ 21.7	0.15	0.45	0.657	1.25(0.17),1.32(0.15) 1.27(0.16),1.31(0.17)
LambdaA	600.1/ 603.5	578.6/ 585.9	0.023	0.484	0.622	0.87(0.1),0.90(0.09) 0.901(0.07),0.885(0.095)
Lambda1A	623.5/ 624.8	612.8/ 611.2	0.28	0.372	0.604	0.914(0.057),0.923(0.052) 0.908(0.057),0.926(0.052)
Lambda2A	595.9/ 600.1	575.1/ 581.7	0.017	0.554	0.726	0.807(0.062),0.837(0.061) 0.826(0.068),0.827(0.072)
ThetaA	2578/ 2575	2603/ 2589	0.536	0.487	0.648	154.1(41),153.6(41.2) 154.9(43.1),148.2(38.4)
Adj.GammaA	689.6/ 689.2	675.4/ 664.3	0.137	0.463	0.685	0.812(0.08),0.825(0.07) 0.846(0.07),0.817(0.08)
Width_parentA	276.3/ 280.7	294.9/ 301.2	0.015	0.786	0.882	3.93(0.71),3.81(0.69) 3.86(0.73),3.79(0.65)
Width_child1A	208.3/ 210.9	226.2/ 228.2	0.05	0.81	0.863	3.65(0.61),3.57(0.62) 3.56(0.6),3.57(0.6)
Width_child2A	170.4/ 168.2	195.3/ 182.4	0.29	0.667	0.763	3.16(0.43),3.19(0.47) 3.197(0.46),3.125(0.43)
Width_allA	144.5/ 146.2	157.9/ 159.8	0.043	0.645	0.875	3.585(0.528),3.524(0.559) 3.544(0.565),3.493(0.51)
Angle.BC_A	2531/ 2529	2556/ 2543	0.265	0.754	0.708	127.4(41.7),121.1(40.5) 126.9(38.7),120.2(35.5)
AlphaV	427.4/ 436.6	404.7/ 409.2	0.001	0.33	0.577	0.68(0.19),0.71(0.19) 0.737(0.21),0.725(0.19)
BetaV	203.5/ 206.7	174.8/ 185.5	0.026	0.344	0.62	1.11(0.18),1.15(0.19) 1.14(0.19),1.16(0.21)
LambdaV	1123/ 1133	1100/ 1105	0.002	0.338	0.593	0.815(0.12),0.832(0.12) 0.85(0.13),0.843(0.12)
Lambda1V	1677/ 1672	1657/ 1640	0.35	0.335	0.574	0.906(0.062),0.914(0.067) 0.902(0.072),0.909(0.065)
Lambda2V	1371/ 1381	1349/ 1352	<0.000	0.362	0.636	0.733(0.087),0.754(0.082) 0.759(0.081),0.761(0.091)
ThetaV	182.4/ 177.8	214.2/ 196	0.71	0.353	0.608	137.1(32.9),136.3(28.8) 137.1(34.5),134.7(33)
Adj.GammaV	1929/ 1927	1909/ 1897	0.06	0.351	0.611	0.763(0.067),0.765(0.071) 0.768(0.071),0.761(0.073)
Width_parentV	1282/ 1285	1312/ 1316	0.045	0.624	0.812	4.38(0.9),4.29(0.91) 4.33(0.92),4.21(0.89)
Width_child1V	1078/ 1080	1117/ 1119	0.046	0.62	0.805	4(0.83),3.935(0.75) 3.931(0.792),3.84(0.746)
Width_child2V	734/ 734	766/ 765	0.125	0.521	0.708	3.198(0.531),3.219(0.513) 3.277(0.574),3.22(0.475)
Width_allV	840.5/ 843.6	856.8/ 858.9	0.047	0.333	0.823	3.861(0.694),3.815(0.667) 3.848(0.698),3.792(0.653)
Angle.BC_V	6494/ 6498	6510/ 6515	0.05	0.501	0.611	124.53(38.32),121.61(38.37) 122.96(42.74),119.91(38.15)
Fractal	531.1/ 533.6	515.5/ 521.3	0.047	0.796	0.844	1.662(0.06),1.648(0.05) 1.647(0.05),1.632(0.06)
Lacunarity	474.3/ 469.8	465.3/ 451.8	0.688	0.79	0.832	0.16(0.082),0.17(0.073) 0.171(0.059),0.174(0.078)
CRAE_LEON	608.3/ 611.8	626.5/ 629.3	0.01	0.75	0.818	20.12(3.04),19.37(2.93) 19.52(3.03),19.28(2.75)
CRVE_LEON	736.3/ 745.4	753.5/ 756.8	0.001	0.745	0.826	28.88(4.77),26.88(4.4) 27.28(5.03),27.04(4.73)
AVR_LEON	294.5/ 289.9	285.6/ 272.2	0.31	0.445	0.549	0.691(0.081),0.712(0.084) 0.718(0.091),0.723(0.092)

CRAE_KNUD	589.7/ 595.8	604.5/ 607.4	0.007	0.776	0.843	20.10(2.99),19.33(2.86) 19.56(3.07),19.87(2.82)
CRVE_KNUD	744.3/ 753.4	761.1/ 763.1	0.001	0.737	0.821	29.56(4.93),27.50(4.45) 28.19(5.03),27.14(4.89)
AVR_KNUD	302.1/ 301.1	294.1/ 283.9	0.1	0.443	0.562	0.688(0.081),0.704(0.082) 0.702(0.094),0.723(0.093)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value

Table A.2: Analysis of tortuosity features for study 1

Feature name	AIC ^a	BIC ^a	(p- χ^2) ^b	ICC	Ω^2	Group Means(SD) (Y3, Y2, Y1, DR)
mean_phi	2494/ 2489	2481/ 2464	0.665	0.538	0.631	0.222(0.03),0.223(0.024) 0.223(0.026),0.225(0.028)
median_phi	2222/ 2218	2209/ 2193	0.488	0.551	0.644	0.139(0.032),0.138(0.033) 0.14(0.037),0.143(0.035)
75th_phi	1739/ 1734	1726/ 1708	0.823	0.491	0.589	0.347(0.052),0.35(0.04) 0.346(0.052),0.35(0.056)
sd_phi	2758/ 2754	2746/ 2728	0.72	0.339	0.447	0.24(0.019),0.242(0.017) 0.24(0.018),0.241(0.018)
mean_tau	3550/ 3546	3537/ 3520	0.6544	0.506	0.603	(3.71(0.98),3.76(0.85) 3.72(0.99),3.8(0.93))x10 ⁻²
median_tau	4687/ 4684	4674/ 4658	0.371	0.652	0.728	(11.1(3.6),11(3.1) 10.9(3.7),11.3(3.7))x10 ⁻³
75th_tau	3641/ 3637	3628/ 3611	0.593	0.554	0.645	(35.4(9),35.2(7.8) 35.1(9.2),36(9.4))x10 ⁻³
sd_tau	2452/ 2455	2422/ 2425	0.05	0.295	0.406	(7.99(2.4),8.3(2.3) 8.1(2.5),8.4(2.3))x10⁻²
mean_psi	3633/ 3629	3621/ 3603	0.575	0.539	0.632	(4.47(0.92),4.49(0.86) 4.45(0.89),4.55(0.88))x10 ⁻²
median_psi	4276/ 4272	4263/ 4246	0.493	0.708	0.772	(16.9(5.7),16.7(5.2) 16.9(5.8),17.3(5.4))x10 ⁻³
75th_psi	3254/ 3250	3241/ 3224	0.716	0.606	0.687	(51.4(13.6),51.9(12.7) 52.1(13.3),52.6(13.9))x10 ⁻³
sd_psi	3104/ 3098	3091/ 3073	0.389	0.275	0.383	(7.25(1.3),7.25(1.2) 7.2(1.3),7.3(1.3))x10 ⁻²
Tort.blend	989/ 986	977/ 961	0.41	0.576	0.664	1.009(0.114),1.016(0.103) 1.012(0.114),1.024(0.122)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value

Table A.3: Post-Hoc analysis for study 1

Significant LMM features	Significant Linear Hypotheses	P-values
AlphaA	y3-y2	0.045
Lambda	y3-y2	0.04
Lambda2A	y3-y2	<0.000
Width_allA	y3-DR	0.032
	y2-DR	0.043
Width_parentA	y3-DR	0.032
	y3-y2	0.045
Width_child1A	y3-DR	0.048
AlphaV	y3-DR	0.019
	y3-y1	0.001
BetaV	y3-DR	0.035
Lambda	y3-DR	0.017
	y3-y1	0.002

Lambda2	y3-DR	0.001
	y3-y1	0.003
	y3-y2	0.03
Width_allV	y3-DR	0.012
	y2-DR	0.048
	y3-y1	0.029
Width_parentV	y3-DR	0.049
Width_child1V	y3-DR	0.048
Angle_BC_V	y3-DR	0.032
Fractal	y3-DR	0.026
sd_tau	y3-DR	0.035
CRAE_LEON	y3-DR	0.017
	y1-DR	0.049
CRVE_LEON	y3-DR	0.034
	y2-DR	0.05
	y1-DR	0.012
CRAE_KNUD	y3-DR	0.023
	y2-DR	0.026
	y1-DR	0.039
CRVE_KNUD	y3-DR	0.036
	y2-DR	0.021
	y1-DR	0.029

Appendix A.2 Four year Non-Matched Study of the Progressors

Table A.4: Analysis of the main features for study 2

Feature name	AIC ^a	BIC ^a	(p-(χ^2) ^b	ICC	Ω^2	Group Means(SD) (Y3, Y2, Y1, DR)
AlphaA	505.3/	486.7/	0.498	0.388	0.324	0.801(0.136),0.82(0.145)
	501.6	468.9				0.808(0.125),0.807(0.127)
BetaA	192.3/	226.4/	0.846	0.567	0.247	1.302(0.193),1.311(0.202)
	187.2	206.4				1.323(0.197),1.303(0.209)
LambdaA	1509/	1490/	0.461	0.391	0.341	0.858(0.103),0.899(0.065)
	1505	1472				0.883(0.042),0.881(0.054)
Lambda1A	1867/	1848/	0.71	0.645	0.241	0.922(0.068),0.92(0.071)
	1862	1828				0.926(0.062),0.921(0.065)
Lambda2A	1570/	1551/	0.61	0.419	0.319	0.813(0.079),0.822(0.091)
	1566	1532				0.822(0.091),0.815(0.093)
ThetaA	1156/	1189/	0.108	0.687	0.369	145.6(45.1),152.4(49.7)
	1155	1175				149.5(48.5),153.9(46.1)
Adj.GammaA	2070/	2051/	0.865	0.574	0.246	0.844(0.079),0.841(0.079)
	2065	2031				0.85(0.074), 0.828(0.079)
Width_allA	1384/	1417/	0.159	0.236	0.609	3.422(0.619),3.534(0.643)
	1383	1402				3.496(0.652),3.486(0.625)
Width_parentA	1948/	1985/	0.474	0.284	0.559	3.727(0.821),3.841(0.839)
	1942	1967				3.791(0.848),3.796(0.814)
Width_child1A	1582/	1621/	0.236	0.229	0.628	3.488(0.714),3.583(0.726)
	1580	1617				3.561(0.729),3.549(0.729)
Width_child2A	1049/	1072/	0.02	0.234	0.504	3.049(0.475),3.178(0.504)
	1052	1076				3.137(0.501),3.112(0.481)
Angle_BC_A	9216/	9250/	0.444	0.77	0.347	116.40(43.81),119.1(47.39)
	9213	9232				117.19(45.5),121.42(46.33)
AlphaV	934.2/	910.8/	0.198	0.273	0.175	0.713(0.209),0.728(0.198)
	932.3	891.9				0.735(0.205),0.734(0.207)
BetaV	413.2/	454.1/	0.136	0.497	0.215	1.201(0.208),1.217(0.257)
	412.8	436.1				1.232(0.212),1.219(0.203)
LambdaV	3421/	3397/	0.188	0.248	0.181	0.834(0.129),0.844(0.121)
	3419	3378				0.848(0.125),0.847(0.127)
Lambda1V	5629/	5605/	0.758	0.591	0.15	0.915(0.067),0.916(0.068)
	5623	5583				0.919(0.061),0.915(0.063)

Lambda2V	3750/ 3749	3725/ 3709	0.075	0.323	0.223	0.757(0.102),0.768(0.092) 0.773(0.101),0.769(0.104)
ThetaV	2328/ 2327	2369/ 2351	0.129	0.108	0.104	137.2(39.1),138.6(38.4) 137.1(38.5),133.6(38.6)
Adj.GammaV	5990/ 5989	5967/ 5948	0.179	0.521	0.21	0.778(0.056),0.783(0.061) 0.787(0.063),0.773(0.059)
Width_allV	4698/ 4697	4738/ 4722	0.074	0.125	0.421	3.587(0.71),3.636(0.685) 3.671(0.706),3.638(0.692)
Width_parentV	6310/ 6306	6351/ 6330	0.581	0.149	0.392	4.009(0.956),4.042(0.924) 4.07(0.941), 4.047(0.943)
Width_child1V	5645/ 5642	5686/ 5665	0.336	0.122	0.375	3.723(0.84),3.759(0.797) 3.794(0.82),3.759(0.836)
Width_child2V	2981/ 3003	3022/ 3027	<0.000	0.123	0.407	3.031(0.494),3.108(0.475) 3.149(0.521),3.108(0.466)
Angle.BC_V	24684/ 24687	24705/ 24711	0.045	0.3	0.154	117.49(37.93),117.38(36.67) 115.74(36.81),114.17(35.21)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value

Table A.5: Post-Hoc analysis for study 2

Significant LMM features	Significant Linear Hypotheses	P-values
Width_child2A	y3-y1	0.042
	y3-y2	0.033
Width_child2V	y3-DR	0.003
	y3-y1	<0.000
	y3-y2	0.004
Angle.BC_V	y3-DR	0.039

Appendix A.3 Four Year Matched Study of the Non-Progressors

Table A.6: Analysis of the main features for study 3

Feature name	AIC ^a	BIC ^a	(p-(χ^2) ^b	ICC	Ω^2	Group Means(SD) (Y3, Y2, Y1, Y0)
AlphaA	60.7/ 61.1	50.7/ 43.5	0.092	0.2	0.145	0.740(0.163),0.805(0.171), 0.82(0.161),0.842(0.142)
BetaA	5.99/ 4.76	12.81/ 4.05	0.189	0.04	0.341	1.259(0.23),1.35(0.234), 1.38(0.24),1.321(0.212)
LambdaA	169.8/ 168.1	158.2/ 151.2	0.088	0.277	0.457	0.855(0.095),0.872(0.056), 0.891(0.069),0.895(0.044)
Lambda1A	202.6/ 197.9	192.5/ 180.4	0.138	0.346	0.334	0.846(0.079),0.856(0.082), 0.863(0.081),0.839(0.077)
Lambda2A	193/ 190.1	180.1/ 175.4	0.231	0.188	0.431	0.806(0.083),0.819(0.083), 0.815(0.061),0.821(0.055)
ThetaA	69.39/ 67.55	86.97/ 67.55	0.244	0.234	0.4	143.7(34.1),143.9(33.1), 144.9(33.2),158.6(40.6)
Adj.GammaA	236.9/ 235.1	226.9/ 217.6	0.241	0.05	0.334	0.813(0.055),0.824(0.056), 0.835(0.053),0.819(0.051)
Width_parentA	105.1/ 104.4	121.9/ 115.1	0.085	0.467	0.589	4.074(0.505),3.85(0.523), 3.838(0.457),3.912(0.402)
Width_child1A	93.58/ 92.28	111.1/ 102.3	0.195	0.496	0.585	3.834(0.413),3.669(0.454), 3.689(0.474),3.669(0.421)
Width_child2A	89/ 83.71	106.6/ 93.75	0.871	0.373	0.442	3.277(0.465),3.254(0.382), 3.304(0.371),3.325(0.437)
Width_allA	69.23/ 66.48	86.81/ 76.52	0.355	0.505	0.58	3.73(0.415),3.593(0.403), 3.611(0.387),3.635(0.364)

Angle.BC_A	875.6/ 872.5	893.2/ 882.5	0.411	0.211	0.26	112.42(34.42),112.34(37.64), 110.94(35.01),123.99(38.99)
AlphaV	96.49/ 92.91	83.44/ 70.08	0.489	0.464	0.596	0.581(0.206),0.622(0.234), 0.631(0.218),0.613(0.218)
BetaV	43.05/ 41.93	29.99/ 19.1	0.181	0.251	0.433	1.121(0.112),1.173(0.124), 1.164(0.142),1.147(0.116)
LambdaV	267.6/ 263.9	254.6/ 241.1	0.523	0.436	0.57	0.751(0.132),0.775(0.145), 0.782(0.137),0.771(0.141)
Lambda1V	435.3/ 430.5	422.2/ 407.7	0.745	0.206	0.536	0.816(0.057),0.831(0.053), 0.825(0.076),0.822(0.054)
Lambda2V	344.9/ 343.8	331.9/ 320.9	0.186	0.397	0.419	0.681(0.098),0.715(0.109), 0.718(0.114),0.705(0.109)
ThetaV	9.89/ 5.56	32.73/ 18.61	0.642	0.499	0.623	133.1(31.9),135.7(31.6), 136.5(31.4),131.4(30.4)
Adj.GammaV	514.6/ 513.4	501.5/ 490.3	0.211	0.195	0.435	0.758(0.056),0.765(0.043), 0.767(0.056),0.761(0.054)
Width_parentV	394.2/ 391.2	417.8/ 404.3	0.51	0.58	0.611	4.879(0.765),4.722(0.791), 4.844(0.853),4.818(0.827)
Width_child1V	416.9/ 411.6	439.7/ 411.6	0.854	0.41	0.542	4.456(0.747),4.379(0.733), 4.467(0.861),4.435(0.818)
Width_child2V	194.4/ 191.6	217.2/ 204.6	0.36	0.559	0.534	3.284(0.471),3.332(0.505), 3.415(0.499),3.331(0.401)
Width_allV	284.2/ 279.8	307.1/ 292.8	0.668	0.559	0.598	4.206(0.571),4.144(0.588), 4.242(0.646),4.195(0.587)
Angle.BC_V	1792/ 1788	1815/ 1801	0.625	0.422	0.551	123.6(36.37),122.1(32.8), 123.8(33.9),121.4(35.3)
Fractal	146.3/ 144.1	141.1/ 133.4	0.293	0.641	0.732	1.665(0.049),1.651(0.057), 1.64(0.059),1.643(0.055)
Lacunarity	127.5/ 126.9	122.1/ 116.2	0.139	0.564	0.69	0.149(0.031),0.167(0.078), 0.186(0.075),0.181(0.059)
CRAE_LEON	458.98/ 450.49	460.13/ 464.79	0.059	0.053	0.165	18.94(3.19),18.74(3.13), 18.18(3.89),18.01(4.63)
CRVE_LEON	488.93/ 483.23	497.29/ 495.52	0.098	0.067	0.132	25.76(5.11),26.14(3.98), 26.72(5.23),26.08(5.01)
AVR_LEON	18.94/ 15.34	20.09/ 14.56	0.668	<0.000	0.019	0.735(0.321),0.716(0.234), 0.681(0.236),0.691(0.2)
CRAE_KNUD	409.26/ 407.51	421.4/ 416.4	0.052	<0.000	0.118	18.46(2.01),18.24(2.82), 18.02(3.28),17.89(4.32)
CRVE_KNUD	542.38/ 538.81	553.1/ 549.53	0.084	0.045	0.112	25.45(5.01), 26.01(3.78), 26.31(5.11),25.87(4.87)
AVR_KNUD	55.11/ 53.16	46.05/ 40.82	0.067	0.056	0.093	0.731(0.311),0.701(0.245), 0.684(0.245), 0.691(0.201)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value

Table A.7: Analysis of the tortuosity features for study 3

Feature name	AIC ^a	BIC ^a	(p-(χ^2)) ^b	ICC	Ω^2	Group Means(SD) (Y3, Y2, Y1, Y0)
mean_phi	432.19/ 427.91	424.62/ 412.78	0.631	0.327	0.439	0.218(0.021),0.216(0.022), 0.223(0.027),0.218(0.023)
median_phi	376.63/ 373.97	369.06/ 358.83	0.342	0.383	0.506	0.133(0.029),0.133(0.03), 0.145(0.036),0.138(0.035)
75th_phi	304.87/ 300.66	297.3/ 285.53	0.615	0.285	0.397	0.33(0.044),0.345(0.052), 0.345(0.052),0.34(0.044)
sd_phi	487.59/ 483.77	480.02/ 468.64	0.535	0.166	0.268	0.237(0.012),0.236(0.018), 0.24(0.018),0.234(0.019)
mean_tau	617.68/ 612.13	610.11/ 597	0.929	0.518	0.606	0.035(0.007),0.034(0.01), 0.035(0.01),0.035(0.01)
median_tau	822.32/ 821.96	814.4/ 807.19	0.095	0.439	0.576	0.011(0.003),0.0097(0.0025), 0.0112(0.0034),0.0109(0.0033)
75th_tau	671.16/ 667.92	663.6/ 652.79	0.431	0.545	0.643	0.0331(0.0056),0.0317(0.0065), 0.0332(0.0082),0.0340(0.0084)

sd_tau	406.01/ 400.21	398.44/ 385.07	0.978	0.296	0.393	0.081(0.021),0.078(0.031), 0.078(0.034),0.079(0.026)
mean_psi	652.47/ 651.47	644.91/ 636.34	0.172	0.549	0.658	0.044(0.005),0.041(0.008), 0.044(0.009),0.044(0.008)
median_psi	748.19/ 750.84	735.7/ 740.6	0.034	0.44	0.592	0.0158(0.0034),0.0152(0.0038), 0.0178(0.0053),0.0168(0.0048)
75th_psi	616.6/ 615.69	609.04/ 600.56	0.179	0.585	0.687	0.0516(0.008),0.0488(0.0098), 0.0528(0.0127),0.0519(0.0091)
sd_psi	552.11/ 548.14	544.54/ 533.01	0.566	0.407	0.517	0.0724(0.0086),0.0691(0.0136), 0.0721(0.0145),0.0728(0.0141)
Tort.blend	161.55/ 157.43	153.98/ 142.4	0.612	0.375	0.486	0.988(0.091),0.974(0.1), 1.007(0.12),0.986(0.112)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value

Table A.8: Post-Hoc analysis for study 3

Significant LMM features	Significant Linear Hypotheses	P-values
median_psi	y2-y1	0.0325

Appendix A.4 Four Year Non-Matched Study of the Non-Progressors

Table A.9: Analysis of the main features for study 4

Feature name	AIC ^a	BIC ^a	(p-(χ^2) ^b	ICC	Ω^2	Group Means(SD) (Y3, Y2, Y1, Y0)
AlphaA	163.4/ 160	147.8/ 133.44	0.326	0.109	0.152	0.767(0.181),0.763(0.206), 0.804(0.185),0.798(0.181)
BetaA	62.87/ 59.81	90.44/ 75.56	0.401	0.015	0.01	1.285(0.202),1.345(0.214), 1.312(0.198),1.321(0.178)
LambdaA	556.65/ 553.78	541.65/ 526.4	0.377	0.11	0.175	0.869(0.106),0.864(0.127), 0.878(0.109),0.875(0.101)
Lambda1A	828.76/ 827.5	813.24/ 799.87	0.23	0.015	0.069	0.896(0.086),0.866(0.074), 0.845(0.084),0.85(0.071)
Lambda2A	600.9/ 596.9	585.2/ 569.4	0.576	0.064	0.09	0.819(0.103),0.832(0.121), 0.826(0.109),0.827(0.104)
ThetaA	436.16/ 432.63	463.72/ 448.38	0.48	0.09	0.133	142.8(40.1),140.26(38.57), 149.4(47.4),149.6(47.6)
Adj.GammaA	909.68/ 907.21	893.45/ 879.65	0.316	<0.000	0.01	0.823(0.045),0.825(0.056), 0.833(0.052),0.838(0.061)
Width_parentA	765.28/ 763.34	793.88/ 778.83	0.373	0.17	0.25	3.823(0.677),3.958(0.710), 4.004(0.717),3.96(0.623)
Width_child1A	735.51/ 735.27	762.83/ 751.25	0.101	0.173	0.322	3.599(0.619),3.814(0.712), 3.766(0.668),3.762(0.659)
Width_child2A	550.21/ 546.42	573.98/ 565.75	0.121	0.151	0.233	3.100(0.535),3.244(0.527), 3.297(0.463),3.285(0.496)
Width_allA	602.17/ 601.77	629.33/ 618.22	0.082	0.189	0.302	3.508(0.56),3.672(0.578), 3.689(0.55),3.669(0.53)
Angle.BC_A	288.99/ 285.55	316.55/ 301.28	0.467	0.046	0.071	115.34(39.33),111.12(35.21), 119.23(44.53),117.87(40.04)
AlphaV	157.69/ 154.94	144.53/ 128.62	0.355	0.159	0.179	0.647(0.222),0.668(0.223), 0.677(0.221),0.645(0.216)
BetaV	36.65/ 31.33	62.92/ 44.94	0.877	0.054	0.066	1.141(0.145),1.159(0.122), 1.154(0.156),1.134(0.175)
LambdaV	706.57/ 703.82	693.4/ 677.5	0.354	0.165	0.185	0.792(0.141),0.805(0.138), 0.811(0.14),0.791(0.139)
Lambda1V	1343/ 1337	1330/ 1311	0.927	0.045	0.062	0.836(0.058),0.836(0.059), 0.831(0.062),0.833(0.063)

Lambda2V	847.72/ 844.22	834.56/ 817.9	0.475	0.146	0.158	0.737(0.122),0.751(0.118), 0.751(0.122),0.734(0.127)
ThetaV	411.26/ 409.12	437.58/ 422.28	0.277	0.129	0.152	129.9(33.9),133.3(38.2), 134.9(35.1),127.7(32.6)
Adj.GammaV	1474/ 1469	1461/ 1442	0.875	0.04	0.05	0.754(0.069),0.767(0.056), 0.774(0.034),0.773(0.063)
Width_parentV	1555/ 1551	1582/ 1565	0.515	0.197	0.201	4.456(0.97),4.436(0.875), 4.541(0.91),4.63(1.001)
Width_child1V	1465/ 1461	1491/ 1474	0.679	0.2	0.208	4.156(0.901),4.141(0.809), 4.216(0.865),4.299(0.933)
Width_child2V	808.06/ 805.76	404.03/ 402.88	0.221	0.188	0.216	3.204(0.523),3.262(0.517), 3.332(0.498),3.303(0.474)
Width_allV	1200/ 1197	1226/ 1210	0.349	0.208	0.215	3.939(0.732),3.946(0.667), 4.031(0.678),4.077(0.731)
Angle.BC_V	286.56/ 281.81	312.88/ 294.97	0.741	0.093	0.114	115.93(36.15),118.12(36.92), 118.88(36.34),116.4(35.49)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value

Appendix A.5 Progressors Versus Non-Progressors Versus DR

Table A.10: Analysis of the main features for study 5

Feature name	AIC ^a	BIC ^a	(p-(χ^2) ^b	ICC	Ω^2	Group Means(SD) (Prog, Non-Prog, DR)
AlphaA	234.19/ 230.28	218.91/ 208.15	0.352	0.161	0.215	0.75(0.212),0.785(0.203) 0.792(0.184)
BetaA	133.85/ 135.7	151.89/ 148.98	0.054	0.031	0.735	1.253(0.171),1.317(0.162) 1.27(0.176)
LambdaA	881.51/ 879.54	868.23/ 857.40	0.362	0.169	0.225	0.836(0.117),0.878(0.101) 0.872(0.109)
Lambda1A	1306/ 1304	1292/ 1282	0.372	0.072	0.109	0.912(0.057),0.922(0.052) 0.927(0.044)
Lambda2A	950.77/ 950.37	480.39/ 478.18	0.111	0.112	0.154	0.803(0.116),0.832(0.111) 0.824(0.113)
ThetaA	806.5/ 804.62	826.75/ 819.78	0.053	0.108	0.151	147.61(50.47),143.73(43.48) 158.77(48.56)
Adj.GammaA	1429/ 1427	1414/ 1407	0.058	0.031	0.742	0.782(0.068),0.800(0.063) 0.788(0.068)
Width_parentA	1471/ 1479	1501/ 1504	0.002	0.341	0.957	4.413(1.071),4.062(0.671) 3.937(0.856)
Width_child1A	1357/ 1362	1376/ 1379	0.008	0.373	0.409	3.932(0.914),3.663(0.648) 3.675(0.795)
Width_child2A	908.11/ 915.12	928.39/ 930.24	0.004	0.271	0.778	3.461(0.641),3.167(0.453) 3.181(0.51)
Width_allA	1155/ 1162	1171/ 1174	0.003	0.373	0.402	4.002(0.839),3.564(0.524) 3.597(0.669)
Angle.BC_A	6329/ 6328	6350/ 6343	0.076	0.112	0.148	122.17(49.54),114.42(39.72) 125.73(44.77)
AlphaV	49.06/ 47.76	78.98/ 76.54	0.16	0.576	0.703	0.672(0.231),0.635(0.233) 0.699(0.234)
BetaV	95.45/ 91.12	207.83/ 204.34	0.23	0.558	0.688	1.141(0.163),1.182(0.141) 1.175(0.178)
LambdaV	677/54/ 674.34	663.45/ 660.34	0.601	0.141	0.099	0.806(0.149),0.797(0.155) 0.822(0.148)
Lambda1V	1525/ 1522	1512/ 1499	0.867	0.048	0.077	0.908(0.069),0.912(0.066) 0.912(0.074)
Lambda2V	853.44/ 850.97	839.82/ 828.28	0.465	0.123	0.165	0.724(0.123),0.717(0.109) 0.722(0.125)
ThetaV	577.06/ 576.57	599.26/ 590.68	0.105	0.098	0.132	135.91(37.66),125.55(32.26) 137.7(39.7)

Adj.GammaV	1661/ 1659	1647/ 1636	0.475	0.09	0.123	0.753(0.064),0.753(0.055) 0.762(0.068)
Width_parentV	2031/ 2034	2047/ 2053	0.03	0.344	0.943	4.83(1.215),4.423(0.896) 4.476(1.154)
Width_child1V	1925/ 1928	1941/ 1948	0.035	0.276	0.3	4.458(1.099),4.116(0.893) 4.147(1.061)
Width_child2V	1171/ 1175	1188/ 1194	0.021	0.276	0.778	3.482(0.691),3.173(0.431) 3.294(0.585)
Width_allV	1618/ 1622	1636/ 1641	0.021	0.346	0.345	4.256(0.925),3.904(0.652) 3.972(0.861)
Angle.BC_V	7845/ 7865	7912/ 7918	0.001	0.321	0.506	119.56(37.87),121.45(38.56) 117.62(36.08)
Fractal	628.32/ 625.48	618.53/ 609.16	0.559	0.837	0.759	1.652(0.058),1.65(0.054), 1.637(0.064)
Lacunarity	543.34/ 539.6	533.56/ 523.29	0.878	0.806	0.717	0.167(0.071),0.171(0.063), 0.176(0.078)
CRAE_LEON	1025/ 1036	1042/ 1046	<0.000	0.506	0.593	19.57(3.13),16.29(4.08) 19.51(2.75)
CRVE_LEON	1163/ 1165	1175/ 1177	0.05	0.515	0.607	27.71(4.92),24.78(5.14) 27.69(4.74)
AVR_LEON	182.2/ 179.9	172.4/ 163.4	0.451	0.077	0.131	0.682(0.098),0.715(0.092), 0.713(0.094)
CRAE_KNUD	955.9/ 964.2	972.3/ 974.8	0.002	0.595	0.677	19.58(3.03),17.01(3.18) 19.98(2.82)
CRVE_KNUD	1245/ 1248	1255/ 1257	0.016	0.413	0.511	28.38(5.03),24.61(7.09) 27.99(4.78)
AVR_KNUD	293.4/ 291.9	283.5/ 275.4	0.291	0.092	0.165	0.698(0.088),0.726(0.159), 0.723(0.093)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value

Table A.11: Analysis of the tortuosity features for study 5

Feature name	AIC ^a	BIC ^a	(p-(χ^2)) ^b	ICC	Ω^2	Group Means(SD) (Prog, Non-Prog, DR)
mean_phi	925.1/ 924.6	915.1/ 907.9	0.197	0.434	0.532	0.229(0.029),0.218(0.023), 0.227(0.027)
median_phi	799.7/ 797.5	789.1/ 780.9	0.406	0.487	0.584	0.147(0.043),0.135(0.032), 0.143(0.037)
75th_phi	650.1/ 649.9	640.3/ 632.7	0.213	0.395	0.495	0.361(0.056),0.338(0.046), 0.356(0.053)
sd_phi	1062/ 1065	1048/ 1052	0.035	0.144	0.198	0.246(0.017),0.237(0.016) 0.249(0.018)
mean_tau	1335/ 1336	1325/ 1319	0.107	0.382	0.503	0.041(0.011),0.036(0.009), 0.037(0.009)
median_tau	1780/ 1778	1770/ 1762	0.284	0.405	0.512	0.012(0.003),0.011(0.003), 0.012(0.004)
75th_tau	650.2/ 649.2	640.2/ 632.2	0.063	0.361	0.484	0.037(0.008),0.033(0.007), 0.036(0.009)
sd_tau	916.8/ 915.1	906.8/ 898.5	0.313	0.305	0.415	0.089(0.026),0.081(0.027), 0.083(0.024)
mean_psi	1364/ 1363	1354/ 1347	0.202	0.374	0.488	0.047(0.010),0.044(0.008), 0.046(0.009)
median_psi	1586/ 1584	1576/ 1568	0.297	0.542	0.645	0.018(0.007),0.016(0.005), 0.018(0.005)
75th_psi	1225/ 1222	1215/ 1205	0.669	0.453	0.549	0.055(0.015),0.051(0.011), 0.053(0.013)
sd_psi	1173/ 1170	1163/ 1154	0.638	0.226	0.329	0.075(0.015),0.072(0.012), 0.073(0.014)
Tort.blend	333.4/ 334.9	322.1/ 323.3	0.061	0.379	0.481	1.056(0.113),0.991(0.105), 1.031(0.115)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value

Table A.12: Post-Hoc analysis for study 5

Significant LMM features	Significant Linear Hypotheses	P-values
Width_parentA	prog-nonprog	0.001
	prog-DR	0.014
Width_child1A	prog-nonprog	0.011
	prog-DR	0.021
Width_child2A	prog-nonprog	0.006
	prog-DR	0.013
Width_allA	prog-nonprog	0.003
	prog-DR	0.011
Width_parentV	prog-DR	0.042
Width_child1V	prog-DR	0.046
Width_child2V	prog-DR	0.043
Width_allV	prog-DR	0.03
Angle.BC_V	prog-nonprog	<0.000
	nonprog-DR	<0.000
CRAE_LEON	prog-nonprog	<0.000
	nonprog-DR	<0.000
CRVE_LEON	prog-nonprog	0.04
CRAE_KNUD	prog-nonprog	<0.002
	nonprog-DR	<0.001
CRVE_KNUD	prog-nonprog	0.01
sd_phi	prog-nonprog	0.021

Appendix A.6 Progressors Four Year Study - Independent Groups

Table A.13: Analysis of the main features for study 6

Feature name	AIC ^a	BIC ^a	(p-(χ^2) ^b	ICC	Ω^2	Group Means(SD) (Y3, Y2, Y1, DR)
AlphaA	115.86/ 113.24	102.97/ 90.66	0.336	<0.000	0.417	0.779(0.184),0.831(0.165), 0.791(0.145),0.813(0.156)
BetaA	125.04/ 120.06	148.76/ 133.62	0.795	0.489	0.804	1.265(0.301),1.25(0.324), 1.33(0.256),1.351(0.231)
LambdaA	101.83/ 96.01	88.27/ 72.92	0.978	0.789	0.965	0.807(0.103),0.818(0.106), 0.812(0.122),0.805(0.105)
Lambda1A	428.52/ 424.15	414.97/ 400.43	0.653	0.04	0.586	0.902(0.067),0.902(0.061), 0.922(0.059),0.926(0.061)
Lambda2A	364.38/ 360.76	350.82/ 337.03	0.497	0.164	0.865	0.815(0.106),0.851(0.095), 0.835(0.081),0.841(0.097)
ThetaA	301.68/ 297.93	325.41/ 311.49	0.523	0.316	0.431	1.236(0.351),1.247(0.371), 1.176(0.367),1.347(0.387)
Adj.GammaA	22.78/ 17.04	46.52/ 30.61	0.967	0.326	0.992	0.838(0.093),0.842(0.097), 0.839(0.097),0.829(0.095)
Width_parentA	587.68/ 581.75	611.4/ 595.31	0.995	0.752	0.835	3.408(1.346),3.255(1.55), 3.507(1.318),3.383(1.139)
Width_child1A	441.67/ 436.48	465.4/ 450.04	0.847	0.435	0.991	3.506(0.764),3.613(0.724), 3.638(0.77),3.431(0.604)
Width_child2A	318.86/ 315.35	342.45/ 328.91	0.476	0.446	0.738	3.072(0.516),3.3(0.609), 3.241(0.589),3.08(0.416)
Width_allA	414.74/ 409.17	438.46/ 422.72	0.934	0.602	0.902	3.329(0.742),3.39(0.812), 3.462(0.792),3.297(0.608)
Angle.BC_A	2284/ 2281	2307/ 2294	0.432	0.044	0.116	128.64(54.54),129.45(50.65), 122.45(42.54),128.78(48.32)
AlphaV	88.68/ 88.56	74.56/ 64.93	0.111	0.014	0.659	0.732(0.212),0.713(0.197), 0.686(0.186),0.771(0.179)

BetaV	42.39/ 42.27	66.05/ 55.98	0.106	0.05	0.459	1.136(0.136),1.116(0.153), 1.081(0.182),1.166(0.174)
LambdaV	307.79/ 307.71	294.11/ 284	0.107	0.023	0.662	0.846(0.125),0.835(0.123), 0.821(0.117),0.872(0.105)
Lambda1V	480.63/ 476.89	467.03/ 453.11	0.519	<0.000	0.479	0.912(0.061),0.91(0.063), 0.892(0.086),0.908(0.075)
Lambda2V	326.45/ 327.62	303.12/ 304.56	0.067	0.076	0.576	0.794(0.111),0.783(0.117), 0.753(0.116),0.816(0.105)
ThetaV	219.88/ 220.56	234.45/ 235.65	0.056	0.045	0.218	145.75(43.5),136.4(36.18), 133.68(40.64),125.77(33.95)
Adj.GammaV	507.39/ 507.3	493.34/ 492.65	0.115	0.051	0.428	0.779(0.065),0.773(0.07), 0.753(0.077),0.783(0.068)
Width_parentV	595.66/ 593.48	619.45/ 607.07	0.281	0.307	0.345	3.873(0.906),4.155(1.091), 4.306(0.996),3.783(0.718)
Width_child1V	515.45/ 513.08	539.24/ 526.68	0.304	0.332	0.369	3.63(0.8),3.871(0.924), 3.931(0.831),3.516(0.534)
Width_child2V	339.97/ 334.89	363.76/ 348.48	0.82	0.293	0.322	3.025(0.644),3.157(0.544), 3.155(0.468),3.034(0.396)
Width_allV	461.69/ 458.37	485.47/ 471.96	0.443	0.362	0.385	3.509(0.74),3.727(0.81), 3.797(0.706),3.444(0.502)
Angle.BC_V	2190/ 2188	2212/ 2204	0.056	0.053	0.430	119.33(37.67),116.57(38.96), 122.01(45.67),108.87(35.23)
Fractal	103.45/ 99.19	98.51/ 89.06	0.656	0.837	0.891	1.676(0.085),1.669(0.031), 1.644(0.077),1.642(0.053)
Lacunarity	110.34/ 109.43	85.12/ 83.43	0.345	0.745	0.456	0.163(0.131),0.156(0.102), 0.177(0.098),0.164(0.065)

^a: Full versus Restricted (null) model - the lower the better, ^b: chi-square p-value